Error estimates for the gradient discretisation of degenerate parabolic equation of porous medium type

Clément Cancès, Jerome Droniou, Cindy Guichard, Gianmarco Manzini, Manuela Bastidas Olivares, Iuliu Sorin Pop

ERROR ESTIMATES FOR THE GRADIENT DISCRETISATION OF DEGENERATE PARABOLIC EQUATION OF POROUS MEDIUM TYPE

CLÉMENT CANCÈS, JÉRÔME DRONIOU, CINDY GUICHARD, GIANMARCO MANZINI, MANUELA BASTIDAS OLIVARES, AND IULIU SORIN POP

ABSTRACT. The gradient discretisation method (GDM) is a generic framework for the spatial discretisation of partial differential equations. The goal of this contribution is to establish an error estimate for a class of degenerate parabolic problems, obtained under very mild regularity assumptions on the exact solution. Our study covers well-known models like the porous medium equation and the fast diffusion equations, as well as the strongly degenerate Stefan problem. Several schemes are then compared in a last section devoted to numerical results.

1. INTRODUCTION

Degenerate parabolic equations appear as mathematical models for numerous real-life applications, like reactive solute transport in porous media, water infiltration in the vadose zone, geological CO_2 sequestration, oil recovery, biological systems, or phase transition problems. In the simplest form, one has

$$\begin{aligned} \partial_t \overline{u} - \Delta \zeta(\overline{u}) &= f && \text{in } (0,T)\times\Omega, \\ \zeta(\overline{u}) &= 0 && \text{on } (0,T)\times\partial\Omega, \\ \overline{u}(0,\cdot) &= u_{\rm ini} && \text{on } \Omega. \end{aligned} \tag{1}$$

With L^∞ denoting the space of essentially bounded functions and $\|\cdot\|_\infty$ the corresponding norm, throughout this chapter we assume the following.

(A0) $T > 0$ and Ω is a bounded connected domain of $\mathbb{R}^d$ ($d \in \mathbb{N}^*$) with Lipschitz continuous boundary $\partial\Omega$.
(A1) $\zeta : \mathbb{R} \to \mathbb{R}$ is continuous, non-decreasing and satistfies $\zeta(0) = 0$.
(A2) $u_{\rm ini} \in L^\infty(\Omega)$ with $M_0 := \|u_{\rm ini}\|_\infty$.
(A3) $f \in L^\infty((0,T]\times\Omega)$ with $M_f := \|f\|_\infty$.

As follows from (A1), ζ' may become zero, or unbounded for certain arguments $\overline{u}$. Consequently, the equation may degenerate from a parabolic equation into an elliptic or an ordinary one. The degeneracy regions are not known a-priori, but depend on the solution itself and may change in time.

One of the most representative example in this sense, the porous medium equation (PME), appeared in the last century as a mathematical model for the flow of an ideal gas in a porous medium. In this case one has

$$\zeta(\overline{u}) = |\overline{u}|^{m-1}\overline{u} \text{ for some } m > 1. \tag{2}$$

Compared to the heat equation, which is obtained for $m = 1$ and in which the equation is linear and parabolic everywhere regardless of the data, if $m > 1$ the nonlinear diffusive term vanishes if $\overline{u} \le 0$, and the equation degenerates. In particular, this leads to the occurrence of free boundaries separating regions in Ω where $\overline{u} > 0$ from those where $\overline{u} \le 0$. These free boundaries have an a-priori unknown location and move in time with a finite speed, which is the reason for calling such cases as "slow diffusion" ones.

Another remarkable example in the category of "slow diffusion" equations is the Stefan problem, which models phase transition problems like melting or solidification. In this case

$$\zeta(\overline{u}) = \begin{cases} \overline{u}, & \text{if } u < 0, \\ \max\{0, \overline{u} - 1\}, & \text{if } u \ge 1. \end{cases} \tag{3}$$

Though bounded, ζ' is vanishing on the entire interval $(0,1)$.

A different situation appears when ζ is as in (2), but with $m \in (0,1)$. In this case no free boundaries occur, but $\zeta' \to \infty$ whenever $\overline{u} \to 0$ so the diffusion coefficient becomes unbounded.

This equation is also known as generalized porous medium equation (GPME), and one speaks about a "fast diffusion". It can appear as a mathematical model for reactive transport in porous media, for equilibrium kinetics (see [6]).

The degeneracy has direct impact on the regularity of the solutions. Unlike the regular parabolic case, the solutions to degenerate parabolic problems have lower regularity, and the singularities are not smoothed out but may even develop in time. Such effects are particularly encountered at the free boundaries. The lack of regularity motivates the introduction of a notion of weak solution.

We use standard notations and function spaces in the functional analysis: $L^2(\Omega)$, $L^\infty(\Omega)$, $H^1_0(\Omega)$, or its dual $H^{-1}(\Omega)$. Whenever obvious, the domain Ω is left out. With X being one of the spaces before, $L^2(0,T;X)$ is the space of X-valued measurable functions that are square integrable in the sense of Bochner. We let $(\cdot,\cdot)$ stand for the inner product on $L^2(\Omega)$, or the duality pairing between $H^1_0(\Omega)$ and $H^{-1}(\Omega)$, and $\|\cdot\|$ for the norm in $L^2(\Omega)$, or the straightforward extension to $L^2(\Omega)^d$, and $\|\cdot\|_\infty$ is the L^∞ norm in Ω or in $(0,T]\times\Omega$. We will often write u or $u(t)$ instead of $u(t,\boldsymbol{x})$ and use C to denote a generic positive constant independent of the discretisation parameters or the function itself.

We start by defining a weak solution for (1):

Definition 1.1. *A weak solution to* (1) *is a function* $\overline{u}\in H^1(0,T;H^{-1}(\Omega))$ *s.t.* $\zeta(\overline{u})\in L^2(0,T;H^1_0(\Omega))$, $\overline{u}(0)=u_{\rm ini}$ *in* $H^{-1}(\Omega)$ *and for a.e.* $t\in(0,T]$ *and* $v\in H^1_0(\Omega)$ *it holds*

$$(\partial_t\overline{u}(t),v)+(\nabla\zeta(\overline{u}(t)),\nabla v)=(f(t),v). \tag{4}$$

The existence and uniqueness of a weak solution to (1) is proved e.g. in [2] and [42] in the case where ζ is increasing. If ζ is merely nondecreasing, existence and uniqueness still hold, see e.g. [15], as well as [44]. As already suggested, the degenerate aspect of the problem makes the usual regularity theory for parabolic problems (see for instance [35]) fail. What is kept is mainly the following:

- *Maximum principle:* the solution $\overline{u}$ belongs to $L^\infty((0,T]\times\Omega)$, with

$$\|\overline{u}\|_\infty\le M_0+TM_f. \tag{5}$$

- *Energy estimate:* Consider the primitive of ζ defined by $\Xi:\mathbb{R}\to\mathbb{R}$, $\Xi(v)=\int_0^v\zeta(z)\mathrm{d}z$. Ξ is convex and positive and one has

$$\int_\Omega\Xi(\overline{u}(t))+\frac12\int_0^t\int_\Omega|\nabla\zeta(s)|^2\le\int_\Omega\Xi(u_{\rm ini})+\frac12\|f\|^2_{L^2(0,T;H^{-1}(\Omega))}. \tag{6}$$

- *Continuity of* $\zeta(\overline{u})$: it is shown in [54] under quite general assumptions on ζ (including cases where ζ is constant on an interval) that $\zeta(\overline{u})$ belongs to $C((0,T]\times\Omega)$. In the case where ζ is increasing (thus invertible), one gets that $\overline{u}\in C((0,T]\times\Omega)$ too. Because of the degeneracy of the problem, this estimate is not enough to initiate a bootstrap to recover the usual parabolic regularity theory.
- *Time continuity of* $\overline{u}$: even if ζ is not invertible, one can still give a (weaker) sense to $\overline{u}(t)$ as a function (and not only as a distribution in H^{-1} as suggested by Definition 1.1). Indeed, $\overline{u}\in C([0,T];L^p(\Omega))$ for all $p\in[1,+\infty)$ thanks to [11].

Further regularity results in the PME case where $\zeta(u)=|u|^{m-1}u$ (or more generally when ζ is increasing) can be found in the monographs [49, 50] (see also [36] for the local Hölder continuity), while the Stefan problem is extensively discussed in [39].

The literature on the numerical approximation of degenerate parabolic equations is extremely rich. Often, the numerical scheme are combining a regularisation step, that is used to deal with the lack of regularity of the solution to degenerate problems. Whenever regularisation is involved, this is mostly obtained through a perturbation ζ_ε of ζ, of which derivative is bounded away from 0 from below and from infinity from above (see e.g. [41]). Alternatively, one can exploit the maximum principle and perturb the boundary and initial data in such a way that the solution stays away from values at which degeneracy is encountered.

Concerning various specific numerical schemes, we mention that often the time stepping is of first order. In particular Euler implicit of semi-implicit methods are popular, and this is due to the lack in regularity of the solution. For the spatial discretisation we mention the conformal finite element schemes analysed e.g. in [41] for the slow diffusion, or in [6] for the fast diffusion. The convergence of the mixed finite element discretisation is proved in [5, 52] for the slow diffusion case, and for

a range allowing for both kind of degeneracies in [48]. We also mention [53] for the analysis of a scheme combining mortars with mixed finite elements. These papers are proving the convergence of the scheme by obtaining a-priori error estimates rigorously. The convergence of finite volume schemes is proved in [3, 4, 29, 31] by means of compactness arguments, and in [30] for a finite volume phase-by-phase upstream weighting. Error estimates are obtained in [34] for a multipoint flux approximation scheme by using the equivalence with a mixed finite element scheme, and in [45] for the simplest two-point approximation in the slow dffusion case, but under minimal regularity assumptions. Discontinuous Galerkin schemes for porous media flow models leading to degenerate parabolic equations are analysed e.g. in [26, 27]. To conclude this paragraph, we mention that a posteriori error estimates for degenerate problems related to porous media flows are derived in [14], [51].

The goal of this chapter is to study in a general way a large class of numerical approximation of (1), entering the framework of the so-called *Gradient Discretisation Method (GDM)* [23]. This general framework is detailed in Section 3. The ideas used in the numerical analysis below apply to methods which are energetically stable (i.e., a discrete counterparts of (6) holds). Our approach does not require any monotonicity properties on the approximation, like the maximum principle. This choice is due to the fact that proving the maximum principle (5), as well as the (time-)continuity for $\zeta(\overline{u})$ and $\overline{u}$, relies strongly on the monotonicity of equation (1), which also extends to the time-discrete case but not to the fully discrete provided by the GDM. We mention that the convergence of the GDM for a class of problems covering (1) is obtained in [28, 21] by means of compactness arguments. The aim here is to extend such results by providing a-prori error estimates

Remark 1.2. *The present results can be adapted without any particular further difficulty to the case of problems with anisotropy of the form*

$$\partial_t \overline{u} - \nabla \cdot (\Lambda \nabla \zeta(\overline{u})) = f \qquad \text{in } (0,T)\times\Omega,$$

where $\Lambda \in L^\infty(\Omega;\mathbb{R}^{d\times d})$ is a symmetric definite positive tensor field, i.e., $\Lambda(\boldsymbol{x}) = \Lambda(\boldsymbol{x})^T$ and there exists $\lambda_m, \lambda_M > 0$ such that

$$\lambda_m |v|^2 \le \Lambda(\boldsymbol{x}) v \cdot v \le \lambda_M |v|^2, \qquad \boldsymbol{x} \in \Omega,\ v \in \mathbb{R}^d.$$

The paper is organised as follows. Section 2 is introducing the sequence of semi-discrete in time problems. The time discretisation relies on the backward Euler scheme and is thus very standard. The a-priori error estimates for the time discretisation are deeply inspired from [38], and do not require any regularity assumption on the exact solution. Section 3 is devoted to the fully discrete setting. This encompasses the definition of the notions of *Gradient Discretisation* and *Gradient Scheme*, which were introduced in [25] and further developed in the monograph [23]. The main result is an error estimate for any scheme entering this general framework of the GDM. To this purpose, reasonable extra regularity slightly overpassing the aforementioned regularity results rigorously established in the literature will be assumed on the solution $\overline{u}$. Finally, several numerical schemes are compared in Section 4; these schemes consist of the Locally Enriched Non-Conforming Polytopal scheme, the Hybrid Mimetic Mixed method, two versions of the Vertex Approximate Gradient scheme, the mass-lumped $\mathbb{P}^1$ Finite Element scheme, the Hybridizable Discontinuous Galerkin scheme, and the Conforming Virtual Element Method.

2. Semi-discrete in time problem

Our purpose in this section is to show how to derive an error estimate using only minimal regularity assumptions for time-discrete approximations of (1). To this end, we first establish some a-priori estimates on the time-discrete solution. This section shall be seen as a first step towards the derivation of the fully discrete error estimate of Theorem 3.4.

2.1. The time discretisation.

In view of the low regularity of the solution, we only consider first order time discretisation schemes. To this aim we consider a sequence of times $0 = t^{(0)} < t^{(1)} < \cdots < t^{(N)} = T$ ($N \in \mathbb{N}^*$) and define the time steps $\delta t^{(n+\frac{1}{2})} = t^{(n+1)} - t^{(n)}$ ($n \in \{0,\dots,N-1\}$). We let $\overline{u}^{(n)}$ be a time discrete approximation of $\overline{u}(t^{(n)})$. To define a weak solution to the time-discrete problems we use the set $X_0 := \{\Xi(u) \in L^1(\Omega) \,:\, \zeta(u) \in H_0^1(\Omega)\}$. The Euler implicit discretisation of (4) consists in finding a sequence of solutions to the time discrete problems, as defined in

Definition 2.1 (Time semi-discrete problem)**.** *Set* $\overline{u}^{(0)} = u_{\rm ini}$*. With* $n \in \{0, \dots, N-1\}$*, given* $\overline{u}^{(n)} \in X_0$*, a weak solution* $\overline{u}^{(n+1)} \in X_0$ *to the time discrete problem at time step* $t^{(n+1)}$ *satisfies, for all* $v \in H_0^1(\Omega)$*,*

$$(\overline{u}^{(n+1)}, v) + \delta t^{(n+\frac{1}{2})}(\nabla\zeta(\overline{u}^{(n+1)}), \nabla v) = (\overline{u}^{(n)}, v) + \delta t^{(n+\frac{1}{2})}(f^{(n+1)}, v), \tag{7}$$

where $f^{(n+1)}(\boldsymbol{x}) = \frac{1}{\delta t^{(n+\frac{1}{2})}} \int_{t^{(n)}}^{t^{(n+1)}} f(s, \boldsymbol{x})\, ds$*.*

Theorem 2.2 (Existence and uniqueness of a solution to the semi-discrete problem)**.** *There exists a unique family* $(\overline{u}^{(n)})_{n=0,\dots,N}$ *solution to the semi-discrete problem in the sense of Definition 2.1.*

Proof. Follows by applying [22, Theorem A.1] to solve, at each step, the non-linear elliptic problem $w - \delta t^{(n+\frac{1}{2})}\Delta\zeta(w) = \overline{u}^{(n)} + \delta t^{(n+\frac{1}{2})} f^{(n+1)}$. □

2.2. **A-priori estimates.** Our goal is to provide a fully discrete error analysis for numerical schemes for (1). For ease of legibility, we start by discussing some properties of the semi-discrete, Euler implicit discretisation in (7). In doing so, we follow the ideas in [38], where the convergence of a linear, time discrete scheme is proved for a class of problems that includes (1).

We start with a remark on the essential boundedness of a solution. This property is physically justified for many of the applications that can be modelled mathematically in the form of (1) (e.g. the gas flow in porous media flows, or the reactive transport). In this context, the essential boundedness is inherited by the time discrete solutions, which satisfy a maximum principle. However, since this property does not extend to the fully discrete cases excepting some particular finite element or finite volume discretisation, we will avoid using it below.

Assuming that the initial data and the source term are both essentially bounded, as stated in Assumptions (A2) and (A3), one has

Lemma 2.3. *Assume* $\overline{u}^{(n)} \in X_0$ *is such that* $\|\overline{u}^{(n)}\|_\infty \le M_0 + M_f t^{(n)}$*. Then the solution* $\overline{u}^{(n+1)}$ *of* (7) *satisfies* $\|\overline{u}^{(n+1)}\|_\infty \le M_0 + M_f t^{(n+1)}$*.*

These estimates are obtained straightforwardly by testing in (7) with $v = [\zeta(\overline{u}^{(n+1)}) - \zeta(M_0 + M_f t^{(n+1)})]_+$, and with $v = [\zeta(\overline{u}^{(n+1)}) + \zeta(M_0 + M_f t^{(n+1)})]_-$ (with $[s]_+ = \max(s, 0)$ and $[s]_- = \min(s, 0)$). We omit the details.

We state some elementary results that are used below, and which are valid for all set of vectors $\boldsymbol{a}_n, \boldsymbol{b}_n \in \mathbb{R}^d$ $(d \ge 1)$, $n \in \{0, \dots, m\}$.

$$2\sum_{n=1}^{m} \boldsymbol{a}_n \cdot (\boldsymbol{a}_n - \boldsymbol{a}_{n-1}) = |\boldsymbol{a}_m|^2 - |\boldsymbol{a}_0|^2 + \sum_{n=1}^{m} |\boldsymbol{a}_n - \boldsymbol{a}_{n-1}|^2, \tag{8}$$

$$2\sum_{n=0}^{m}\sum_{j=0}^{n} \boldsymbol{a}_n \cdot \boldsymbol{a}_j = \left|\sum_{n=0}^{m} \boldsymbol{a}_n\right|^2 + \sum_{n=0}^{m} |\boldsymbol{a}_n|^2, \tag{9}$$

$$\sum_{n=1}^{m} \boldsymbol{a}_n \cdot (\boldsymbol{b}_n - \boldsymbol{b}_{n-1}) = \boldsymbol{a}_m \cdot \boldsymbol{b}_m - \boldsymbol{a}_0 \cdot \boldsymbol{b}_0 - \sum_{n=1}^{m} (\boldsymbol{a}_n - \boldsymbol{a}_{n-1}) \cdot \boldsymbol{b}_{n-1}. \tag{10}$$

Further, with the convex, positive primitive Ξ of ζ appearing in (6), a classical convexity relation yields

$$(b - a)\zeta(b) \ge \Xi(b) - \Xi(a)\,, \qquad \forall a, b \in \mathbb{R}. \tag{11}$$

Finally, we state for completeness the Young inequality, valid for any $a, b \in \mathbb{R}$ and $\varepsilon > 0$,

$$ab \le \frac{1}{2\varepsilon} a^2 + \frac{\varepsilon}{2} b^2. \tag{12}$$

The stability of the time discrete scheme is stated in

Lemma 2.4. *Let* $(\overline{u}^{(n+1)})_{n=0,\dots,N-1}$ *be the sequence of time discrete solutions introduced in Definition 2.1. Then,*

$$\sum_{n=0}^{N-1} \delta t^{(n+\frac{1}{2})} \|\nabla\zeta(\overline{u}^{(n+1)})\|^2 \le C. \tag{13}$$

Proof. Taking in (7) $v = \zeta(\overline{u}^{(n+1)})$ gives

$$(\overline{u}^{(n+1)} - \overline{u}^{(n)}, \zeta(\overline{u}^{(n+1)})) + \delta t^{(n+\frac{1}{2})} \|\nabla \zeta(\overline{u}^{(n+1)})\|^2 = \delta t^{(n+\frac{1}{2})} (f^{(n+1)}, \zeta(\overline{u}^{(n+1)})). \tag{14}$$

For the first term one uses (11) to obtain

$$(\overline{u}^{(n+1)} - \overline{u}^{(n)}, \zeta(\overline{u}^{(n+1)})) \geq \int_\Omega \Xi(\overline{u}^{(n+1)}) - \Xi(\overline{u}^{(n)}) \, dx. \tag{15}$$

The second term needs no further discussion, whereas for the term on the right one obtains

$$\delta t^{(n+\frac{1}{2})} |(f^{(n+1)}, \zeta(\overline{u}^{(n+1)}))| \leq \frac{\delta t^{(n+\frac{1}{2})}}{2} \|f^{(n+1)}\|^2_{H^{-1}(\Omega)} + \frac{\delta t^{(n+\frac{1}{2})}}{2} \|\nabla \zeta(\overline{u}^{(n+1)})\|^2. \tag{16}$$

Using (15) and (16) into (14) and summing the resulting relation over $n \in \{1, \dots, N-1\}$ yields

$$\int_\Omega \Xi(\overline{u}^{(n+1)}) + \frac{1}{2} \sum_{n=0}^{N-1} \delta t^{(n+\frac{1}{2})} \|\nabla \zeta(\overline{u}^{(n+1)})\|^2 \leq \int_\Omega \Xi(\overline{u}^{(0)}) + \sum_{n=0}^{N-1} \frac{\delta t^{(n+\frac{1}{2})}}{2} \|f^{(n+1)}\|^2_{H^{-1}(\Omega)}. \tag{17}$$

The first term on the right is bounded due to Assumption (A2). For the second term on the right one uses Assumption (A3) and the fact that $L^\infty((0,T) \times \Omega)$ is continuously embedded into $L^2(0,T; H^{-1}(\Omega))$, to obtain a uniform bound w.r.t. N. Since Ξ is a positive function, the first term on the left is positive, which provides (13). □

Remark 2.5 (Other estimates)**.** *Other a-priori estimates can be obtained if further assumptions are made on ζ and on the initial data. For example, if ζ is Lipschitz and $\zeta(u_{\text{ini}}) \in H^1_0(\Omega)$, then one can prove that $\|\nabla \zeta(\overline{u})\|$ and $\|\nabla \zeta(\overline{u}^{(n)})\|$ are uniformly bounded (w.r.t. t, respectively n) and obtain L^2 estimates for $\partial_t \zeta(\overline{u})$ or its time discrete counterpart.*

2.3. **Error estimates, time discrete case.** We now establish error estimates for the time discrete scheme. The proof follows the lines of [38, Section 4]. We use the following notations for the errors

$$\begin{aligned} e_u(t) &:= \overline{u}(t) - \overline{u}^{(n+1)}, \\ e_\zeta(t) &:= \zeta(\overline{u}(t)) - \zeta(\overline{u}^{(n+1)}), \\ e_\gamma(t) &:= \gamma(\overline{u}(t)) - \gamma(\overline{u}^{(n+1)}), \end{aligned}$$

for $t \in (t^{(n)}, t^{(n+1)}]$ and $n \in \{0, \dots, N-1\}$. With this, one has

Theorem 2.6. *Let $\overline{u}$ be the solution in Definition 1.1, and $(\overline{u}^{(n+1)})_{n=0,\dots,N-1}$ be the sequence of solutions to the time discrete problems in Definition 2.1. Setting $\delta t = \max_{n \in \{0,\dots,N-1\}} \delta t^{(n+\frac{1}{2})}$, one has*

$$\max_{n \in \{0,\dots,N-1\}} \|e_u(t^{(n+1)})\|^2_{H^{-1}(\Omega)} + \int_0^T (e_u(t), e_\zeta(t)) \, dt \leq C \delta t. \tag{18}$$

Remark 2.7. *Although the second term in (18) is not a proper norm, it can generate an error estimate in a certain norm whenever the particular form of ζ is taken into account. This idea is exploited, in the fully discrete setting, in Corollary 3.7 below. Moreover, as for the a-priori estimates, under additional assumptions on ζ one can get L^2 error estimates for either $\zeta(\overline{u})$ (if ζ is Lipschitz continuous, as appearing in the slow diffusion case) or $\overline{u}$ (if ζ is bijective and its inverse Lipschitz, as appearing in the fast diffusion case).*

Proof. Before giving the proof we observe that, due to the monotonicity of ζ, the second term in (18) is positive.

With $j \in \{0, \dots, N-1\}$ we integrate (4) in time for $t \in (t^{(j)}, t^{(j+1)}]$ and subtract (7) for $n = j$ to obtain

$$(e_u(t^{(j+1)}) - e_u(t^{(j)}), v) + \left(\nabla \int_{t^{(j)}}^{t^{(j+1)}} e_\zeta(t) dt, \nabla v \right) = 0,$$

for all $v \in H^1_0(\Omega)$. After summation over $j \in \{0, \dots, p\}$ for some $p \in \{0, \dots, N-1\}$, this yields

$$\left(e_u(t^{(p+1)}), v \right) + \left(\nabla \int_0^{t^{(p+1)}} e_\zeta(t) dt, \nabla v \right) = 0, \tag{19}$$

for all $v \in H_0^1(\Omega)$. Taking $v = \int_{t^{(p)}}^{t^{(p+1)}} e_\zeta(t)dt$ in the above equation provides

$$\left(e_u(t^{(p+1)}), \int_{t^{(p)}}^{t^{(p+1)}} e_\zeta(t)dt\right) + \sum_{j=0}^{p} \left(\nabla \int_{t^{(j)}}^{t^{(j+1)}} e_\zeta(t)dt, \nabla \int_{t^{(p)}}^{t^{(p+1)}} e_\zeta(t)dt\right) = 0. \tag{20}$$

Fixing $n \in \{0, \dots, N-1\}$, we sum (20) over $p \in \{0, \dots, n\}$ and obtain

$$\underbrace{\sum_{p=0}^{n} \left(e_u(t^{(p+1)}), \int_{t^{(p)}}^{t^{(p+1)}} e_\zeta(t)dt\right)}_{=:I_1} + \underbrace{\sum_{p=0}^{n}\sum_{j=0}^{p} \left(\nabla \int_{t^{(j)}}^{t^{(j+1)}} e_\zeta(t)dt, \nabla \int_{t^{(p)}}^{t^{(p+1)}} e_\zeta(t)dt\right)}_{=:I_2} = 0. \tag{21}$$

The first term can be rewritten as

$$I_1 = \underbrace{\int_0^{t^{(n+1)}} \big(e_u(t), e_\zeta(t)\big)dt}_{=:I_{11}} + \underbrace{\sum_{p=0}^{n} \int_{t^{(p)}}^{t^{(p+1)}} \Big(\overline{u}(t^{(p+1)}) - \overline{u}(t), e_\zeta(t)\Big)\, dt}_{=:I_{12}}. \tag{22}$$

Being positive, I_{11} needs no further handling. For I_{12} we write

$$\overline{u}(t^{(p+1)}) - \overline{u}(t) = \int_t^{t^{(p+1)}} \partial_s \overline{u} ds$$

to obtain

$$\begin{aligned} |I_{12}| &= \left| \sum_{p=0}^{n} \int_{t^{(p)}}^{t^{(p+1)}} \left(\int_t^{t^{(p+1)}} \partial_s \overline{u} ds, e_\zeta(t) \right) dt \right| \\ &\le \left(\max_{n \in \{0,\dots,N-1\}} \delta t^{(n+\frac{1}{2})} \right) \|\partial_t \overline{u}\|_{L^2(0,T;H^{-1}(\Omega))} \|\nabla e_\zeta\|_{L^2(0,T;L^2(\Omega))}. \end{aligned}$$

The regularity of the weak solution prescribed in Definition 1.1 and the a-priori estimate (13) ensure the existence of a $C > 0$ not depending on the time discretisation so that

$$\|\partial_t \overline{u}\|_{L^2(0,T;H^{-1}(\Omega))} \le C, \qquad \|\nabla e_\zeta\|_{L^2(0,T;L^2(\Omega))} \le C.$$

As a consequence, we obtain that

$$|I_{12}| \le C\delta t. \tag{23}$$

Finally, using (9), I_2 is nonnegative and can be underestimated by

$$I_2 \ge \frac{1}{2} \left\| \nabla \int_0^{t^{(n+1)}} e_\zeta(t)dt \right\|^2. \tag{24}$$

Since n was chosen arbitrarily, using (22)–(24) in (21) yields

$$\int_0^T \big(e_u(t), e_\zeta(t)\big)dt + \max_{n \in \{0,\dots,N-1\}} \left\| \nabla \int_0^{t^{(n+1)}} e_\zeta(t)dt \right\|^2 \le C\delta t. \tag{25}$$

To complete the proof of Theorem 2.6 one needs to estimate $\|e_u\|_{H^{-1}(\Omega)}$. This follows straightforwardly from (19). For all $v \in H_0^1(\Omega)$ such that $\|v\|_{H_0^1(\Omega)} \le 1$, the Cauchy–Schwarz inequality yields

$$\left(e_u(t^{(p+1)}), v\right) = -\left(\nabla \int_0^{t^{(p+1)}} e_\zeta(t)dt, \nabla v\right) \le \left\| \nabla \int_0^{t^{(p+1)}} e_\zeta(t)dt \right\|.$$

Taking the supremum over such v, squaring and using (25) we infer $\|e_u(t^{(p+1)})\|_{H^{-1}(\Omega)}^2 \le C\delta t$ and the proof is complete. □

3. Gradient discretisation method and generic error estimate

3.1. **Definition of the gradient scheme.** The principle of the GDM is to replace, in the weak formulation of the problem, the continuous space and differential operators by discrete ones. To this aim a discrete space and function/gradient reconstructions on this space are used. Altogether these form a gradient discretisation (GD) $\mathcal{D}$. In general, very few assumptions are made on the GD [23]. However, to deal with the non-linearity we will need, in a similar way as in [22], to consider *nodal* gradient discretisations with *piecewise constant* reconstructions, which also contain the definition of an interpolator. Therefore, we take $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, I_{\mathcal{D}})$ with

- *(Space)* $X_{\mathcal{D},0} = \{v = (v_i)_{i\in I} \,:\, v_i = 0 \quad \forall I_\partial\}$, where I is a finite set and $I_\partial \subset I$ identifies the boundary degrees of freedom.
- *(Function reconstruction)* Given a partition $(U_i)_{i\in I}$ of Ω, for all $v = (v_i)_{i\in I} \in X_{\mathcal{D},0}$, the reconstructed function $\Pi_{\mathcal{D}} v \in L^\infty(\Omega)$ is defined by

$$\Pi_{\mathcal{D}} v = \sum_{i\in I} v_i \mathbf{1}_{U_i}, \tag{26}$$

  where $\mathbf{1}_{U_i}$ is the characteristic function of U_i.
- *(Gradient reconstruction)* $\nabla_{\mathcal{D}} : X_{\mathcal{D},0} \to L^2(\Omega)^d$ is a linear operator such that $\|\nabla_{\mathcal{D}} \cdot\|$ is a norm on $X_{\mathcal{D},0}$.
- *(Interpolator)* The unknowns represent values at points $(\boldsymbol{x}_i)_{i\in\overline{\Omega}}$, with $\boldsymbol{x}_i \in \overline{U}_i$ for all $i \in I$ and $\boldsymbol{x}_i \in \partial\Omega$ whenever $i \in I_\partial$. With $C_{\mathrm{pw},0}(\Omega)$ the set of piecewise continuous functions on $\overline{\Omega}$ that have a zero limit on $\partial\Omega$, we define the interpolator $I_{\mathcal{D}} : C_{\mathrm{pw},0}(\Omega) \to X_{\mathcal{D},0}$ such that

$$(I_{\mathcal{D}}\phi)_i = \operatorname*{ess\text{-}limsup}_{\boldsymbol{x}\to\boldsymbol{x}_i,\, \boldsymbol{x}\in\overline{U}_i} \phi(\boldsymbol{x}), \qquad \forall\, i \in I, \tag{27}$$

  where

$$\operatorname*{ess\text{-}limsup}_{\boldsymbol{x}\to\boldsymbol{x}_i,\, \boldsymbol{x}\in\overline{U}_i} \phi(\boldsymbol{x}) = \lim_{\epsilon\to 0} \operatorname*{ess\text{-}sup}_{B(\boldsymbol{x}_i,\epsilon)\cap\overline{U}_i} \phi.$$

The fact that the function reconstruction is piecewise constant enables us to define, for $g : \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$ and $v \in X_{\mathcal{D},0}$, the element $g(v) \in X_{\mathcal{D},0}$ by applying g to each nodal value: if $v = (v_i)_{i\in I}$, we set $g(v) = (g(v_i))_{i\in I}$. It then holds

$$\Pi_{\mathcal{D}} g(v) = g(\Pi_{\mathcal{D}} v)\,, \qquad \forall v \in X_{\mathcal{D},0}. \tag{28}$$

The subtle choice for the definition of the interpolator is motivated by the following points. Since our study covers the Stefan problem, whose solution might be discontinuous, there is a real need to define an interpolator that allows for merely piecewise continuous functions. If ϕ is continuous, say $\phi \in C_0(\Omega)$, then

$$\operatorname*{ess\text{-}limsup}_{\boldsymbol{x}\to\boldsymbol{x}_i,\, \boldsymbol{x}\in\overline{U}_i} \phi(\boldsymbol{x}) = \phi(\boldsymbol{x}_i), \quad \forall\, i \in I.$$

Therefore, for any continuous function $g : \mathbb{R} \to \mathbb{R}$ and $\phi \in C_{\mathrm{pw},0}(\Omega)$,

$$I_{\mathcal{D}} g(\phi) = g(I_{\mathcal{D}}\phi). \tag{29}$$

When ϕ is only piecewise continuous, then (29) still holds as soon as g is continuous and nondecreasing.

We can now define the gradient scheme for (1) with implicit time stepping. It is obtained from (7) using the discrete space for trial and test function, and replacing the functions and gradients by the corresponding reconstructions. This gives a sequence of fully discrete, nonlinear algebraic problems, obtained for $n \in \{0, \dots, N-1\}$ and starting with $u^{(0)} = I_{\mathcal{D}} u_{\mathrm{ini}}$.

Problem $\mathbf{P}_{\mathcal{D}}^{(n+1)}$: *Given $u^{(n)} \in X_{\mathcal{D},0}$, find $u^{(n+1)} \in X_{\mathcal{D},0}$ such that*

$$\int_\Omega \Pi_{\mathcal{D}}(u^{(n+1)} - u^{(n)})\Pi_{\mathcal{D}} v + \delta t^{(n+\frac{1}{2})} \int_\Omega \nabla_{\mathcal{D}}\zeta(u^{(n+1)}) \cdot \nabla_{\mathcal{D}} v = \delta t^{(n+\frac{1}{2})} \int_\Omega f^{(n+1)} \Pi_{\mathcal{D}} v \tag{30}$$

for all $v \in X_{\mathcal{D},0}$, where $f^{(n+1)}$ is introduced in Definition 2.1.

Proposition 3.1 (Existence and uniqueness of a solution to the gradient scheme [22, Lemma 2.7])**.** *There exists a solution $u = (u^{(n)})_{n=0,\dots,N}$ to the gradient scheme and, if $u_1, u_2 \in X_{\mathcal{D},0}^{N+1}$ are two solutions to this scheme, then $\zeta(u_1) = \zeta(u_2)$ and $\Pi_{\mathcal{D}} u_1 = \Pi_{\mathcal{D}} u_2$.*

Remark 3.2 (Limit to the uniqueness)**.** *If ζ is strictly increasing, then we have complete uniqueness of the solution: $u_1 = u_2$. However, when ζ has plateau this uniqueness may fail* [22, Remark 2.8].

The accuracy of a GD is measured through three quantities: a discrete Poincaré constant $C_{\mathcal{D}}$ (yielding the coercivity of the method), a measure of the defect of the discrete Stokes formula $W_{\mathcal{D}}$ (associated with the limit-conformity of the method), and a measure of the interpolation error $S_{\mathcal{D}}$ (which, when it tends to zero, yields the consistency of the method). The discrete Poincaré constant is

$$C_{\mathcal{D}} = \max_{v \in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\Pi_{\mathcal{D}} v\|}{\|\nabla_{\mathcal{D}} v\|}. \tag{31}$$

The measure of the defect of the discrete Stokes formula is $W_{\mathcal{D}} : H_{\rm div}(\Omega) \to [0,\infty)$ where, for all $\boldsymbol{\psi} \in H_{\rm div}(\Omega)$ (that is, $\boldsymbol{\psi} \in L^2(\Omega)^d$ and ${\rm div}\boldsymbol{\psi} \in L^2(\Omega)$),

$$W_{\mathcal{D}}(\boldsymbol{\psi}) := \max_{v \in X_{\mathcal{D},0}\setminus\{0\}} \frac{1}{\|\nabla_{\mathcal{D}} v\|} \left| \int_\Omega \Pi_{\mathcal{D}} v {\rm div}\boldsymbol{\psi} + \nabla_{\mathcal{D}} v \cdot \boldsymbol{\psi} \right|. \tag{32}$$

In the GDM, the interpolation error usually involves L^2-error in both function and gradient approximation. However, for time-dependent problems such as (1), it will be more efficient to use a weaker norm for the function approximation. We define the discrete H^{-1}-seminorm by: for $\phi \in L^2(\Omega)$,

$$|\phi|_{\mathcal{D},*} := \max\left\{ \int_\Omega \phi \Pi_{\mathcal{D}} v \,:\, v \in X_{\mathcal{D},0}\,,\ \|\nabla_{\mathcal{D}} v\| \le 1 \right\}. \tag{33}$$

We then set, for $\phi \in C_{{\rm pw},0}(\Omega)$ and $\psi \in C_{{\rm pw},0}(\Omega) \cap H^1_0(\Omega)$,

$$S^{\Pi,*}_{\mathcal{D}}(\phi) = |\Pi_{\mathcal{D}} I_{\mathcal{D}} \phi - \phi|_{\mathcal{D},*} \quad \text{and} \quad S^{\nabla}_{\mathcal{D}}(\psi) = \|\nabla_{\mathcal{D}} I_{\mathcal{D}} \psi - \nabla\psi\|. \tag{34}$$

3.2. **A-priori estimates.** We start with the observation that some properties of the solution to the original problem (1), or its time discrete counterpart, are not preserved by the gradient scheme (30). In particular we refer to the maximum principle (see Lemma 2.3), which in the spatially-continuous case is obtained by testing with a cut-off function. However, in the spatially discrete case the cut-off of an element in the finite dimensional space may not belong to that space any more, since the transition from negative to positive values does not necessarily happen at edges or nodes. For obtaining a priori estimates we therefore restrict to using in (30) test functions that are affine functions of $\zeta(u^{(n+1)})$, but mention that schemes allowing to take nonlinear functions are designed and analysed in [10, 12, 13].

Specifically, we extend in this section the estimates in Lemma 2.4 to the fully discrete case.

Lemma 3.3. *For the sequence of fully discrete solutions of* (30) *it holds*

$$\sum_{n=0}^{N-1} \delta t^{(n+\frac12)} \|\nabla_{\mathcal{D}} \zeta(u^{(n+1)})\|^2 \le C_{\mathcal{D}}^2 \|f\|_{L^2(0,T;L^2(\Omega))} + 2\|\Xi(\Pi_{\mathcal{D}} u^{(0)})\|_{L^1(\Omega)}. \tag{35}$$

Proof. Choosing $v = \zeta(u^{(n+1)})$ in (30) leads to

$$\begin{aligned} &\int_\Omega \Pi_{\mathcal{D}}(u^{(n+1)} - u^{(n)}) \Pi_{\mathcal{D}} \zeta(u^{(n+1)}) \\ &\qquad + \delta t^{(n+\frac12)} \int_\Omega \left| \nabla_{\mathcal{D}} \zeta(u^{(n+1)}) \right|^2 = \delta t^{(n+\frac12)} \int_\Omega f^{(n+1)} \Pi_{\mathcal{D}} \zeta(u^{(n+1)}) \end{aligned} \tag{36}$$

for all $n \in \{0,\dots,N-1\}$. Using (28) and the convexity inequality (11), we have

$$\begin{aligned} \int_\Omega \Pi_{\mathcal{D}}(u^{(n+1)} - u^{(n)}) \Pi_{\mathcal{D}} \zeta(u^{(n+1)}) &= \int_\Omega (\Pi_{\mathcal{D}} u^{(n+1)} - \Pi_{\mathcal{D}} u^{(n)}) \zeta(\Pi_{\mathcal{D}} u^{(n+1)}) \\ &\ge \int_\Omega (\Xi(\Pi_{\mathcal{D}} u^{(n+1)}) - \Xi(\Pi_{\mathcal{D}} u^{(n)})). \end{aligned} \tag{37}$$

The right-hand side of (36) can be estimated thanks to Young and discrete Poincaré inequalities as follows:

$$\begin{aligned} \int_\Omega f^{(n+1)} \Pi_{\mathcal{D}} \zeta(u^{(n+1)}) &\le \frac{C_{\mathcal{D}}^2}{2} \|f^{(n+1)}\|^2 + \frac{1}{2C_{\mathcal{D}}^2} \|\Pi_{\mathcal{D}} \zeta(u^{(n+1)})\|^2 \\ &\le \frac{C_{\mathcal{D}}^2}{2} \|f^{(n+1)}\|^2 + \frac12 \|\nabla_{\mathcal{D}} \zeta(u^{(n+1)})\|^2. \end{aligned} \tag{38}$$

Combining (37) and (38) in (36), summing over $n \in \{0, \dots, N-1\}$, and using $\Xi \geq 0$, the proof of (35) is complete. □

3.3. **Error estimate.** With $I_{\mathcal{D}}^n \overline{u} = I_{\mathcal{D}} \overline{u}(t^{(n)})$ for $n \in \{0, \dots, N\}$, we define the errors in $X_{\mathcal{D},0}$:

$$e_{\mathcal{D},u}^{(n)} := u_{\mathcal{D}}^{(n)} - I_{\mathcal{D}}^n \overline{u},$$
$$e_{\mathcal{D},\zeta}^{(n)} := \zeta(u_{\mathcal{D}}^{(n)}) - I_{\mathcal{D}}^n \zeta(\overline{u}),$$

as well as, for $n \in \{1, \dots, N\}$,

$$\varepsilon_{\mathcal{D},\zeta}^{(n)} := \sum_{p=1}^{n} \delta t^{(p-\frac{1}{2})} e_{\mathcal{D},\zeta}^{(p)}.$$

The error estimates for the fully discrete approximation is stated in the following theorem, whose proof is carried out in Section 3.4.

Theorem 3.4 (GDM error estimate for degenerate parabolic problem)**.** *Assume that the solution $\overline{u}$ of (1) satisfies $\overline{u}(t,\cdot) \in C_{\mathrm{pw},0}(\Omega)$ for all $t \in [0,T]$, $\zeta(\overline{u}) \in C([0,T]; C_0(\Omega) \cap H_0^1(\Omega))$, and $\nabla\zeta(\overline{u}) \in C([0,T]; H_{\mathrm{div}}(\Omega))$. Then, there exists a universal constant K depending neither on the data of the continuous problem nor on the discretisation parameters such that*

$$\max_{1\leq n\leq N} \left|\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)}\right|_{\mathcal{D},*}^2 + \max_{1\leq n\leq N} \left\|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}\right\|^2 + \sum_{n=0}^{N-1} \delta t^{(n+\frac{1}{2})} (\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n+1)}, \Pi_{\mathcal{D}} e_{\mathcal{D},\zeta}^{(n+1)}) \leq K(1+T) E_{\mathcal{D}}(\overline{u})^2, \tag{39}$$

where

$$E_{\mathcal{D}}(\overline{u})^2 = \sum_{n=0}^{N-1} \delta t^{(n+\frac{1}{2})} (E_{\mathcal{D}}^n(\overline{u}))^2 \tag{40}$$

with, for $n \in \{0, \dots, N-1\}$,

$$\begin{aligned} E_{\mathcal{D}}^n(\overline{u}) := {} & \left| \frac{1}{\delta t^{(n+\frac{1}{2})}} \int_{t^{(n)}}^{t^{(n+1)}} \Delta\zeta(\overline{u}(s))\, ds - \Delta\zeta(\overline{u}(t^{(n+1)})) \right|_{\mathcal{D},*} \\ & + S_{\mathcal{D}}^{\Pi,*} \left(\frac{\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})}{\delta t^{(n+\frac{1}{2})}} \right) + S_{\mathcal{D}}^{\nabla}(\zeta(\overline{u}(t^{(n+1)}))) \\ & + W_{\mathcal{D}}(\nabla\zeta(\overline{u}(t^{(n+1)}))). \end{aligned} \tag{41}$$

Due to the non-decreasing property of ζ, each term in the sum in the left-hand side of (39) is non-negative. For the nonlinearities appearing in the case of the Stefan or the porous medium equations, (39) leads to the following error estimates on more natural quantities.

Remark 3.5 (Expected rates of convergence)**.** *Under regularity assumptions on $\overline{u}$, a rate of convergence in terms of time and mesh sizes can be obtained on $E_{\mathcal{D}}(\overline{u})$. Specifically, if $\zeta(\overline{u}) \in C([0,T]; H^2(\Omega))$, $\partial_t \overline{u} \in L^\infty(0,T; H_0^1(\Omega))$ and $\Delta\zeta(\overline{u}) \in W^{1,\infty}(0,T; L^2(\Omega))$, following the techniques in* [23, Section 7.4] *and using $|\phi|_{\mathcal{D},*} \leq C_{\mathcal{D}} \|\phi\|$ it can be proved, for all usual low-order gradient discretisations based on meshes of maximum size h (which include all schemes used in Section 4 except VAG-b), that $E_{\mathcal{D}}^n(\overline{u}) \leq C_{\overline{u}}((1+C_{\mathcal{D}})\delta t^{(n+\frac{1}{2})} + h)$, where $C_{\overline{u}}$ only depends on $\overline{u}$. Hence, in this situation and setting $\delta t = \max_{n\in\{0,\dots,N-1\}} \delta t^{(n+\frac{1}{2})}$, we have*

$$E_{\mathcal{D}}(\overline{u}) \leq T^{\frac{1}{2}} C_{\overline{u}}((1+C_{\mathcal{D}})\delta t + h).$$

Remark 3.6. *For the slow diffusion case and under the regularity stated in Definition 1.1, error estimates for a simple, two-point flux approximation scheme (which fits in the GDM framework) are obtained in* [45]*. The approach there consists in using a discrete Green function to estimate the error for the fully discrete approximation of the sequence of time discrete approximations (Definition 2.1). This approach involves a regularisation step, which we avoid here by using a different strategy.*

Corollary 3.7 (Estimate for the Stefan equation and the PME)**.** *Under the assumptions and notations in Theorem 3.4, the following holds.*

- (Stefan equation) *Assume that ζ is Lipschitz-continuous with Lipschitz constant L_ζ. Then,*

$$\left(\sum_{n=0}^{N-1}\delta t^{(n+\frac12)}\|\zeta(\Pi_{\mathcal D}u^{(n+1)})-\zeta(\Pi_{\mathcal D}I_{\mathcal D}^{n+1}\overline{u})\|^2\right)^{\frac12}\le (K(1+T)L_\zeta)^{\frac12}E_{\mathcal D}(\overline{u}). \tag{42}$$

- (Slow diffusion PME) *Let $\zeta(s)=|s|^{m-1}s$ with $m\ge 1$. There exists $C_m>0$ depending only on m such that*

$$\left(\sum_{n=0}^{N-1}\delta t^{(n+\frac12)}\|\Pi_{\mathcal D}u^{(n+1)}-\Pi_{\mathcal D}I_{\mathcal D}^{n+1}\overline{u}\|_{L^{m+1}(\Omega)}^{m+1}\right)^{\frac{1}{m+1}}\le C_mT^{\frac{1}{m+1}}E_{\mathcal D}(\overline{u})^{\frac{2}{m+1}}. \tag{43}$$

- (Fast diffusion PME) *Let $\zeta(s)=|s|^{m-1}s$ with $m<1$. Then there exists $C_m>0$ depending only on m such that*

$$\left(\sum_{n=0}^{N-1}\delta t^{(n+\frac12)}\|\zeta(\Pi_{\mathcal D}u^{(n+1)})-\zeta(\Pi_{\mathcal D}I_{\mathcal D}^{n+1}\overline{u})\|_{L^{\frac{m+1}{m}}(\Omega)}^{\frac{m+1}{m}}\right)^{\frac{m}{m+1}}\le C_mT^{\frac{m}{m+1}}E_{\mathcal D}(\overline{u})^{\frac{2m}{m+1}}. \tag{44}$$

Corollary 3.7. Stefan equation. Since ζ is a Lipschitz-continuous and non-decreasing function, we have

$$|\zeta(a)-\zeta(b)|^2\le|\zeta(a)-\zeta(b)|L_\zeta|a-b|=L_\zeta(\zeta(a)-\zeta(b))(a-b)\,,\qquad\forall a,b\in\mathbb R.$$

Used in (39), this relation proves (42).

Slow diffusion PME. We first prove that, for some $k_m>0$,

$$|a-b|^{m+1}\le k_m(|a|^{m-1}a-|b|^{m-1}b)(a-b)\,,\quad\forall a,b\in\mathbb R. \tag{45}$$

The case for $b=0$ is trivial and reduces to $k_m\ge1$. Consider $b\neq0$ and set $s=a/b$. To establish (45), we have to prove that $|s-1|^{m+1}\le k_m(|s|^{m-1}s-1)(s-1)$, which reduces to $|s-1|^m\le c_m||s|^{m-1}s-1|$. The function $s\mapsto\frac{|s-1|^m}{||s|^{m-1}s-1|}$ is continuous on $\mathbb R$ (use a Taylor expansion about $s=1$ to deal with the singularity) and has limit 1 at $\pm\infty$. It is therefore bounded, which proves the required estimate.

Using (45) in (39), the estimate (43) follows.

Fast diffusion PME. Let $a',b'\in\mathbb R$ and apply (45) with $\frac1m>1$ instead of m and $a=\zeta(a')=|a'|^{m-1}a'$, $b=\zeta(b')=|b'|^{m-1}b'$. Noting that $|a|^{\frac1m-1}a=a'$ and $|b|^{\frac1m-1}b=b'$, we infer

$$|\zeta(a')-\zeta(b')|^{\frac1m+1}\le k_{1/m}(a'-b')(\zeta(a')-\zeta(b')).$$

Used in (39) this establishes (44). □

3.4. **Proof of Theorem 3.4.** We follow the approach of [18] which consists in identifying an error equation on the discrete solution and the interpolate of the continuous solution, and estimating a consistency error.

To identify the error equation we integrate (4) over the time intervals $(t^{(j)},t^{(j+1)})$ and use (30) to obtain for all $j\in\{0,\dots,N-1\}$

$$\int_\Omega\Pi_{\mathcal D}\left(e_{\mathcal D,u}^{(j+1)}-e_{\mathcal D,u}^{j}\right)\Pi_{\mathcal D}v+\delta t^{(j+\frac12)}\int_\Omega\nabla_{\mathcal D}e_{\mathcal D,\zeta}^{(j+1)}\cdot\nabla_{\mathcal D}v=\mathfrak E_{\mathcal D}^j(v), \tag{46}$$

and all $v\in X_{\mathcal D,0}$, where the consistency error is defined by

$$\mathfrak E_{\mathcal D}^j(v):=\delta t^{(j+\frac12)}\int_\Omega f^{(j+1)}\Pi_{\mathcal D}v-\int_\Omega\left[\Pi_{\mathcal D}I_{\mathcal D}^{j+1}\overline{u}-\Pi_{\mathcal D}I_{\mathcal D}^{j}\overline{u}\right]\Pi_{\mathcal D}v-\delta t^{(j+\frac12)}\int_\Omega\nabla_{\mathcal D}\zeta(I_{\mathcal D}^{j+1}\overline{u})\cdot\nabla_{\mathcal D}v. \tag{47}$$

This is a linear form $\mathfrak E_{\mathcal D}^n:X_{\mathcal D,0}\to\mathbb R$. Its boundedness is established in

Lemma 3.8. *For all $v\in X_{\mathcal D,0}$ and $n\in\{0,\dots,N-1\}$, there holds*

$$|\mathfrak E_{\mathcal D}^n(v)|\le\delta t^{(n+\frac12)}E_{\mathcal D}^n(\overline{u})\|\nabla_{\mathcal D}v\|. \tag{48}$$

Proof. Recalling (34), $\Pi_{\mathcal D}I_{\mathcal D}^{n+1}\overline{u}-\Pi_{\mathcal D}I_{\mathcal D}^{n}\overline{u}$ can be replaced by $\overline{u}(t^{(n+1)})-\overline{u}(t^{(n)})$ in (47) with a cost measured by $S_{\mathcal D}^{\Pi,*}$. Specifically, by the definition of the discrete H^{-1}-seminorm in (33) we have

$$\left|\int_\Omega\phi\Pi_{\mathcal D}v\right|\le|\phi|_{\mathcal D,*}\|\nabla_{\mathcal D}v\|\quad\forall\phi\in L^2(\Omega)\,,\ \forall v\in X_{\mathcal D,0}, \tag{49}$$

and thus, since $I_{\mathcal{D}}^k \overline{u} = I_{\mathcal{D}}(\overline{u}(t^{(k)}))$ for $k = n, n+1$,

$$\begin{aligned}
&\left| \int_\Omega \left[\Pi_{\mathcal{D}} I_{\mathcal{D}}^{n+1}\overline{u} - \Pi_{\mathcal{D}} I_{\mathcal{D}}^{n}\overline{u}\right] \Pi_{\mathcal{D}} v - \int_\Omega \left[\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})\right] \Pi_{\mathcal{D}} v \right| \\
&\quad = \left| \int_\Omega \left[\Pi_{\mathcal{D}} I_{\mathcal{D}} \left(\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})\right) - \left(\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})\right)\right] \Pi_{\mathcal{D}} v \right| \\
&\quad \le S_{\mathcal{D}}^{\Pi,*}(\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})) \|\nabla_{\mathcal{D}} v\|.
\end{aligned}$$

Similarly, replacing $\nabla_{\mathcal{D}}\zeta(I_{\mathcal{D}}^{n+1}\overline{u}) = \nabla_{\mathcal{D}} I_{\mathcal{D}} \zeta(\overline{u}(t^{(n+1)}))$ (see (29)) with $\nabla\zeta(\overline{u}(t^{(n+1)}))$ incurs a cost measured by $S_{\mathcal{D}}^{\nabla}(\zeta(\overline{u}(t^{(n+1)})))$:

$$\left| \int_\Omega \nabla_{\mathcal{D}}\zeta(I_{\mathcal{D}}^{n+1}\overline{u}) \cdot \nabla_{\mathcal{D}} v - \int_\Omega \nabla\zeta(\overline{u}(t^{(n+1)})) \cdot \nabla_{\mathcal{D}} v \right| \le S_{\mathcal{D}}^{\nabla}(\zeta(\overline{u}(t^{(n+1)})))\|\nabla_{\mathcal{D}} v\|.$$

Hence,

$$\begin{aligned}
\mathfrak{E}_{\mathcal{D}}^n(v) = \delta t^{(n+\frac12)} \int_\Omega f^{(n+1)} \Pi_{\mathcal{D}} v - \int_\Omega \left[\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})\right] \Pi_{\mathcal{D}} v - \delta t^{(n+\frac12)} \int_\Omega \nabla\zeta(\overline{u}(t^{(n+1)})) \cdot \nabla_{\mathcal{D}} v \\
+ \mathcal{O}_1\left[S_{\mathcal{D}}^{\Pi,*}(\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})) + \delta t^{(n+\frac12)} S_{\mathcal{D}}^{\nabla}(\zeta(\overline{u}(t^{(n+1)})))\right] \|\nabla_{\mathcal{D}} v\|
\end{aligned}$$

where, here and in the following, $\mathcal{O}_1(X)$ denotes a generic function such that $|\mathcal{O}_1(X)| \le |X|$. Now, by definition of $W_{\mathcal{D}}(\nabla\zeta(\overline{u}(t^{(n+1)})))$,

$$\left| \int_\Omega \nabla\zeta(\overline{u}(t^{(n+1)})) \cdot \nabla_{\mathcal{D}} v + \Delta\zeta(\overline{u}(t^{(n+1)}))\Pi_{\mathcal{D}} v \right| \le W_{\mathcal{D}}(\nabla\zeta(\overline{u}(t^{(n+1)})))\|\nabla_{\mathcal{D}} v\|$$

and thus, writing $\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)}) = \int_{t^{(n)}}^{t^{(n+1)}} \partial_t \overline{u}(s)\, ds$ (which is valid since the equation (1) and the regularity $\nabla\zeta(\overline{u}) \in C([0,T]; H_{\mathrm{div}}(\Omega))$ imply $\partial_t \overline{u} \in C([0,T]; L^2(\Omega))$) and recalling the definition of $f^{(n+1)}$,

$$\begin{aligned}
\mathfrak{E}_{\mathcal{D}}^n(v) = \int_\Omega \left[\int_{t^{(n)}}^{t^{(n+1)}} (f - \partial_t\overline{u})(s)\, ds + \delta t^{(n+\frac12)} \Delta\zeta(\overline{u}(t^{(n+1)}))\right] \Pi_{\mathcal{D}} v \\
+ \mathcal{O}_1\Big[S_{\mathcal{D}}^{\Pi,*}(\overline{u}(t^{(n+1)}) - \overline{u}(t^{(n)})) + \delta t^{(n+\frac12)} S_{\mathcal{D}}^{\nabla}(\zeta(\overline{u}(t^{(n+1)}))) \\
+ \delta t^{(n+\frac12)} W_{\mathcal{D}}(\nabla\zeta(\overline{u}(t^{(n+1)})))\Big] \|\nabla_{\mathcal{D}} v\|.
\end{aligned} \tag{50}$$

Since $f - \partial_t\overline{u} = -\Delta\zeta(\overline{u})$, the property (49) yields

$$\begin{aligned}
&\left| \int_\Omega \left[\int_{t^{(n)}}^{t^{(n+1)}} (f - \partial_t\overline{u})(s)\, ds + \delta t^{(n+\frac12)} \Delta\zeta(\overline{u}(t^{(n+1)}))\right] \Pi_{\mathcal{D}} v \right| \\
&\qquad \le \delta t^{(n+\frac12)} \left| \frac{1}{\delta t^{(n+\frac12)}} \int_{t^{(n)}}^{t^{(n+1)}} \Delta\zeta(\overline{u}(s))\, ds - \Delta\zeta(\overline{u}(t^{(n+1)})) \right|_{\mathcal{D},*} \|\nabla_{\mathcal{D}} v\|.
\end{aligned}$$

Plugging this estimate into (50) and recalling the definition (41) of $E_{\mathcal{D}}^n(v)$, this shows (48). □

With Lemma 3.8 at hand, the next proposition is the main step towards the error estimate in the fully discrete setting.

Proposition 3.9. *For all $n \in \{1, \dots, N\}$, there holds*

$$\max_{1\le n\le N} \|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^n\|^2 + 4 \sum_{n=0}^{N-1} \delta t^{(n+\frac12)} (\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n+1)}, \Pi_{\mathcal{D}} e_{\mathcal{D},\zeta}^{(n+1)}) \le 24T \exp(1) E_{\mathcal{D}}(\overline{u}). \tag{51}$$

Proof. Following the lines of the proof in the semi-discrete case, we sum (46) over $j \in \{0, \dots, p-1\}$ with $p \in \{1, \dots, N\}$, leading to

$$\int_\Omega \Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(p)} \Pi_{\mathcal{D}} v + \sum_{j=0}^{p-1} \delta t^{(j+\frac12)} \int_\Omega \nabla_{\mathcal{D}} e_{\mathcal{D},\zeta}^{(j+1)} \cdot \nabla_{\mathcal{D}} v = \sum_{j=0}^{p-1} \mathfrak{E}_{\mathcal{D}}^j(v). \tag{52}$$

Here, we used the fact that $e^0_{\mathcal{D},u} = 0$ thanks to the definition of $u^0 = I_{\mathcal{D}} u_{\rm ini}$. Choosing $v = \delta t^{(p-\frac12)} e^{(p)}_{\mathcal{D},\zeta}$ in the above equation and summing over $p \in \{1,\dots,n\}$ for some $n \in \{1,\dots,N\}$ provides

$$\mathfrak{I}_1^{(n)} + \mathfrak{I}_2^{(n)} = \mathfrak{R}^{(n)}, \tag{53}$$

where

$$\mathfrak{I}_1^{(n)} := \sum_{p=1}^{n} \delta t^{(p-\frac12)} \int_\Omega \Pi_{\mathcal{D}} e^{(p)}_{\mathcal{D},u} \Pi_{\mathcal{D}} e^{(p)}_{\mathcal{D},\zeta},$$

$$\mathfrak{I}_2^{(n)} := \sum_{p=1}^{n} \sum_{j=1}^{p} \delta t^{(p-\frac12)} \delta t^{(j-\frac12)} \int_\Omega \nabla_{\mathcal{D}} e^{(j)}_{\mathcal{D},\zeta} \cdot \nabla_{\mathcal{D}} e^{(p)}_{\mathcal{D},\zeta},$$

$$\mathfrak{R}^{(n)} := \sum_{p=1}^{n} \sum_{j=0}^{p-1} \delta t^{(p-\frac12)} \mathfrak{E}^j_{\mathcal{D}}(e^{(p)}_{\mathcal{D},\zeta}).$$

$\mathfrak{I}_1^{(n)}$ corresponds to the third term in (39). On the other hand, the identity (9) ensures that

$$\mathfrak{I}_2^{(n)} \geq \frac12 \int_\Omega \left| \nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta} \right|^2. \tag{54}$$

The term $\mathfrak{R}^{(n)}$ can be reorganized as

$$\mathfrak{R}^{(n)} = \sum_{j=0}^{n-1} \mathfrak{E}^j_{\mathcal{D}}(\varepsilon^{(n)}_{\mathcal{D},\zeta}) - \sum_{j=1}^{n-1} \mathfrak{E}^j_{\mathcal{D}}(\varepsilon^{(j)}_{\mathcal{D},\zeta}) := \mathfrak{R}_1^{(n)} + \mathfrak{R}_2^{(n)}.$$

Owing to Lemma 3.8, the first contribution $\mathfrak{R}_1^{(n)}$ can be estimated as follows:

$$\left|\mathfrak{R}_1^{(n)}\right| \leq \sum_{j=0}^{n-1} \left|\mathfrak{E}^j_{\mathcal{D}}(\varepsilon^{(n)}_{\mathcal{D},\zeta})\right| \leq \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} E^j_{\mathcal{D}}(\overline{u}) \|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|.$$

Applying Cauchy–Schwarz inequality and then the Young inequality (12) provides

$$\left|\mathfrak{R}_1^{(n)}\right| \leq T \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} E^j_{\mathcal{D}}(\overline{u})^2 + \frac14 \|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|^2 \leq T E_{\mathcal{D}}(\overline{u})^2 + \frac14 \|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|^2,$$

so that, in view of (54), there holds

$$\mathfrak{I}_2^{(n)} - \mathfrak{R}_1^{(n)} \geq \frac14 \|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|^2 - T E_{\mathcal{D}}(\overline{u})^2. \tag{55}$$

Using once again Lemma 3.8, one can bound the term $\mathfrak{R}_2^{(n)}$ by

$$\mathfrak{R}_2^{(n)} \leq \sum_{j=1}^{n-1} \left|\mathfrak{E}^j_{\mathcal{D}}(\varepsilon^{(j)}_{\mathcal{D},\zeta})\right| \leq \sum_{j=1}^{n-1} \delta t^{(j+\frac12)} E^j_{\mathcal{D}}(\overline{u}) \|\nabla_{\mathcal{D}} \varepsilon^{(j)}_{\mathcal{D},\zeta}\|.$$

Using the Cauchy–Schwarz and Young inequality as above then yields

$$\mathfrak{R}_2^{(n)} \leq 2T E_{\mathcal{D}}(\overline{u})^2 + \frac{1}{8T} \sum_{j=1}^{n-1} \delta t^{(j+\frac12)} \|\nabla_{\mathcal{D}} \varepsilon^{(j)}_{\mathcal{D},\zeta}\|^2. \tag{56}$$

Combining (55)–(56) in (53) provides

$$\|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|^2 + 4 \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} (\Pi_{\mathcal{D}} e^{(j+1)}_{\mathcal{D},u}, \Pi_{\mathcal{D}} e^{(j+1)}_{\mathcal{D},\zeta}) \leq \frac{1}{2T} \sum_{j=1}^{n-1} \delta t^{(j+\frac12)} \|\nabla_{\mathcal{D}} \varepsilon^{(j)}_{\mathcal{D},\zeta}\|^2 + 12 T E_{\mathcal{D}}(\overline{u})^2.$$

The generalized Gronwall Lemma [32, Lemma 5.1] then yields

$$\|\nabla_{\mathcal{D}} \varepsilon^{(n)}_{\mathcal{D},\zeta}\|^2 + 4 \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} (\Pi_{\mathcal{D}} e^{(j+1)}_{\mathcal{D},u}, \Pi_{\mathcal{D}} e^{(j+1)}_{\mathcal{D},\zeta}) \leq 12 T \exp(1) E_{\mathcal{D}}(\overline{u}).$$

Choosing $n = N$, and then $n = \operatorname*{argmax}_j \|\nabla_{\mathcal{D}} \varepsilon^{(j)}_{\mathcal{D},\zeta}\|^2$, we obtain (51). □

With Proposition 3.9 we have estimated $\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)} \Pi_{\mathcal{D}} e_{\mathcal{D},\zeta}^{(n)}$ and $\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}$. The proof of Theorem 3.4 is concluded by establishing the discrete $L^\infty(0,T;H^{-1}(\Omega))$ estimate on $(\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)})_{1\le n\le N}$. This is obtained in the lemma below. Together with Proposition 3.9, this completes the proof of Theorem 3.4.

Lemma 3.10. *For all $n \in \{1,\dots,N\}$, there holds*

$$\left|\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)}\right|_{\mathcal{D},*}^2 \le 2\|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}\|^2 + 2T E_{\mathcal{D}}(\overline{u})^2.$$

Proof. Applying (52) for $p = n$ and with $v \in X_{\mathcal{D},0}$ such that $\|\nabla_{\mathcal{D}} v\| \le 1$, using the Cauchy–Schwarz inequality and recalling (48), we have

$$\int_\Omega \Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)} \Pi_{\mathcal{D}} v \le \|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}\| + \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} E_{\mathcal{D}}^j(\overline{u}).$$

Taking the supremum over such v gives

$$\|\Pi_{\mathcal{D}} e_{\mathcal{D},u}^{(n)}\|_{\mathcal{D},*} \le \|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}\| + \sum_{j=0}^{n-1} \delta t^{(j+\frac12)} E_{\mathcal{D}}^j(\overline{u}) \le \|\nabla_{\mathcal{D}} \varepsilon_{\mathcal{D},\zeta}^{(n)}\| + \sqrt{T} E_{\mathcal{D}}(\overline{u}),$$

and the proof follows straightforwardly. □

4. Numerical examples

4.1. **Numerical results.** For the numerical tests, we consider the porous medium, equation in dimension 2, corresponding to (1) with $\zeta(\overline{u}) = |\overline{u}|^{m-1}\overline{u}$ for $m \in \{2,3,4\}$. The computational domain is given by $T = 1$ and $\Omega = (0,1)^2$, and the exact solution is $\overline{u}(t,x) = \mathcal{B}(t_0 + t, x - x_0)$, where $t_0 = 0.1$, $x_0 = (0.5, 0.5)$ and

$$\mathcal{B}(t,x) = t^{-\frac{1}{m}} \left\{ \left[C_{\mathcal{B}} - \frac{m-1}{4m^2} \left(\frac{|x|}{t^{\frac{1}{2m}}} \right)^2 \right]_+ \right\}^{\frac{1}{m-1}}, \tag{57}$$

is the Barenblatt–Pattle solution. The initial solution is fixed by $u_{\mathrm{ini}} = \overline{u}(0,\cdot)$. We choose $C_{\mathcal{B}} = 0.005$, so that $\mathcal{B}$ remains equal to 0 on $\partial\Omega$ during the entire simulation $t \in [0,1]$. Note that by the offset t_0 in $\mathcal{B}$, the singularity of this function at $t = 0$ is avoided, and the initial condition satisfies Assumption (A2).

The simulations are run over three different mesh families: a family of (mostly) hexagonal meshes, a family of locally refined Cartesian meshes, and a family of triangular meshes. Examples of members of each family are provided in Figure 1. We consider uniform time steps. For the coarsest mesh in each family, the time step is $\delta t^{(n+\frac12)} = 0.1$ for all n; then, for each mesh refinement, the time step is divided by 4. Since we use implicit Euler time stepping, this means that the truncation error in time decay as $\mathcal{O}(h^2)$, where h is the mesh size; as our spatial methods (see below) are low order, in the best possible situation (linear equations, smooth exact solution), the maximal approximation rates are $\mathcal{O}(h)$ on gradients and $\mathcal{O}(h^2)$ on functions. The choice of time steps thus ensures that the spatial truncation error is the leading term in the estimate. The following schemes will be used for the tests.

- LEPNC (Locally Enriched Polytopal Non-Conforming scheme) [24]: applicable on generic polytopal meshes, one unknown per internal edge (after static condensation), based on broken polynomial functions with weak continuity properties across the edges. We have taken a zero weight ϖ on the edge unknowns, so $\Pi_{\mathcal{D}}$ is only computed from the cell unknowns.
- HMM (Hybrid Mimetic Mixed scheme) [23, Chapter 13]: applicable on generic polytopal meshes, one unknown per internal edge (after static condensation), based on local reconstruction of piecewise constant functions and gradients.
- MLP1 (Mass-Lumped $\mathbb{P}^1$ finite element) [23, Section 8.4]: only applicable on triangular meshes, one unknown per vertex, based on standard $\mathbb{P}^1$ shape functions for the gradient and piecewise constant reconstruction around each vertex.
- VAG-a (Vertex Approximate Gradient, first presentation) [23, Section 8.5]: applicable on generic polytopal meshes, one unknown per internal vertex and one unknown per cell, based on standard $\mathbb{P}^1$ on a triangular subdivision of the cells (using the center of the

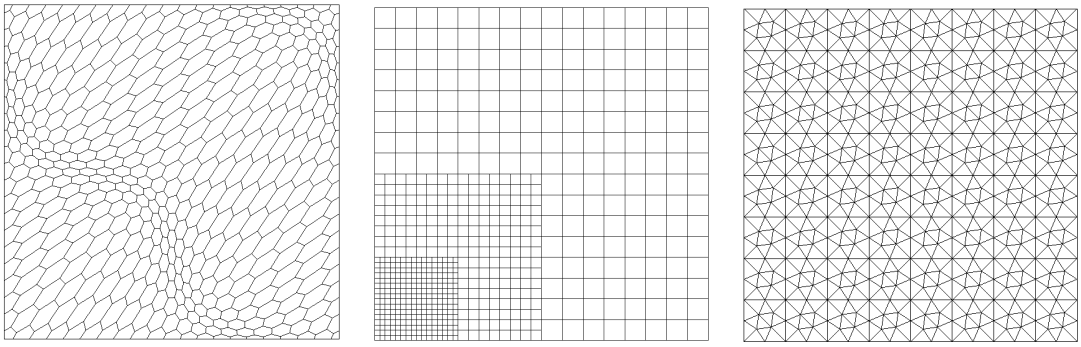


FIGURE 1. Examples of meshes used in numerical tests: hexagonal (left), locally refined Cartesian (centre), and triangular (right).

cell as additional vertex), with a mass-lumping that equally distributes the available area between cell and vertex unknowns. A local algebraic elimination (static condensation) of cell unknowns is also performed, leading to a globally coupled system on the vertex unknowns only.

- VAG-b (Vertex Approximate Gradient, second presentation) [13]: as above, but applied after writing the diffusion term as $\mathrm{div}(m|u|^{m-1}\nabla u)$. Note that this scheme does not present itself as a gradient scheme.
- CFVEM (Conforming Virtual Element) [7]: applicable on generic polytopal meshes, one unknown per internal vertex, based on the elliptic projection of virtual shape functions, with algebraic mass-lumping.
- HDG (Hybridizable Discontinuous Galerkin, order 1) [17]: applicable on generic polytopal meshes (but the results are presented here only on triangular meshes), based on modal Legendre–Dubiner basis functions with one polynomial of degree 1 per cell and edge. The degrees of freedom are reduced to only edge polynomials after static condensation.

The LEPNC and HMM tests are based on the code available from in the `HArDCore2D` library [1] (based on the implementation principles of Hybrid High-Order methods [19]), while MLP1 tests were conducted using the code at https://github.com/jdroniou/matlab-PME.

Remark 4.1 (CFVEM and $\mathbb{P}^1$ finite elements on triangular meshes)**.** *On triangular meshes and for the standard Laplace problem, CFVEM coincides with the conforming $\mathbb{P}^1$ finite element method. The results presented below however show different behaviour of CFVEM and MLP1; the main reason can be found in the mass-lumping strategy adopted for each method: for MLP1, a geometrical mass-lumping was used, allocating to each vertex a mass corresponding to $1/3$ of the sum of the areas of the triangles it belongs to; for CFVEM, an algebraic mass-lumping was used, reducing the standard mass matrix to a diagonal one by summing all elements on each row.*

The accuracy of the schemes are provided through the following quantities, all measured at the final time:

- Relative error in L^2-norm between the (reconstructed) gradients of the approximation of $\zeta(\overline{u})$ and of the interpolate of $\zeta(\overline{u})$:

$$E_{H^1,\zeta} = \frac{\|\nabla_{\mathcal{D}}(\zeta(u^N) - I_{\mathcal{D}}\zeta(\overline{u})(T,\cdot))\|}{\|\nabla_{\mathcal{D}} I_{\mathcal{D}}\zeta(\overline{u})(T,\cdot)\|}. \tag{58}$$

- Relative error in L^{m+1}-norm between the (reconstructed) functions of the approximate solution and of the interpolate of the exact solution $\overline{u}$:

$$E_{L^{m+1}} = \frac{\|\Pi_{\mathcal{D}}(u^N - I_{\mathcal{D}}\overline{u}(T,\cdot))\|_{L^{m+1}(\Omega)}}{\|\Pi_{\mathcal{D}} I_{\mathcal{D}}\overline{u}(T,\cdot)\|_{L^{m+1}(\Omega)}}. \tag{59}$$

- Fraction of negative mass over total mass:

$$\texttt{NMass} = \frac{\displaystyle\int_\Omega (\Pi_{\mathcal{D}} u^N)_-}{\displaystyle\int_\Omega |\Pi_{\mathcal{D}} u^N|}, \tag{60}$$

where $s_- = \max(-s, 0)$ is the negative part of $s \in \mathbb{R}$.

4.1.1. *Rates of convergence vs. mesh size.* We first present the relative errors versus the mesh sizes, for all considered schemes and mesh families. The outputs are given in loglog graphs in Figures 2, 3 and 4. In these figures, the chosen reference slopes correspond to an estimate of the overall behaviour of the schemes, drawn from the tables as well as computed rates of convergence from one mesh to the other. Combining Estimates (39), (43) and Remark 3.5, and considering averaged-in-time norms (which are less stringent than the final time norms (58) and (59)), we would expect for a smooth enough exact solution a rate of convergence $\mathcal{O}(h^{\frac{2}{m+1}})$ in L^{m+1}-norm on $\overline{u}$ and $\mathcal{O}(h)$ in L^2-norm on (the time integral of) $\nabla\zeta(\overline{u})$.

As can be seen in the numerical results, for the considered tests, the theoretical rates for $E_{L^{m+1}}$ are sub-optimal for the lower values of m but tend to be close to the observed results for $m = 4$. The general trend, seen in both the theoretical and numerical results, is that the rates deteriorate as m increases. Two reasons can be found for that: as m increases, the L^{m+1}-norm becomes more constrained, while the regularity of the exact solution $\overline{u}$ decays (for example, it is H^1 in space for $m = 2$, but no longer for $m > 2$).

Interestingly, but not surprisingly, the rates for $E_{H^1,\zeta}$ seem to resist a little bit better as m increases, although they appear to be slightly below 1 for $m > 2$ (which is not surprising since, as mentioned, (58) is a more constraining norm than $\|\nabla_{\mathcal{D}}\varepsilon^{(n)}_{\mathcal{D},\zeta}\|$). Although $E_{H^1,\zeta}$ measures an approximation of the gradient, which can be expected to be of lower order than that of a function, it measures this in a norm that is independent of m and relates to $\zeta(\overline{u})$, a function that has better regularity properties than $\overline{u}$ (for example, it is H^1 in space irrespective of the value of m).

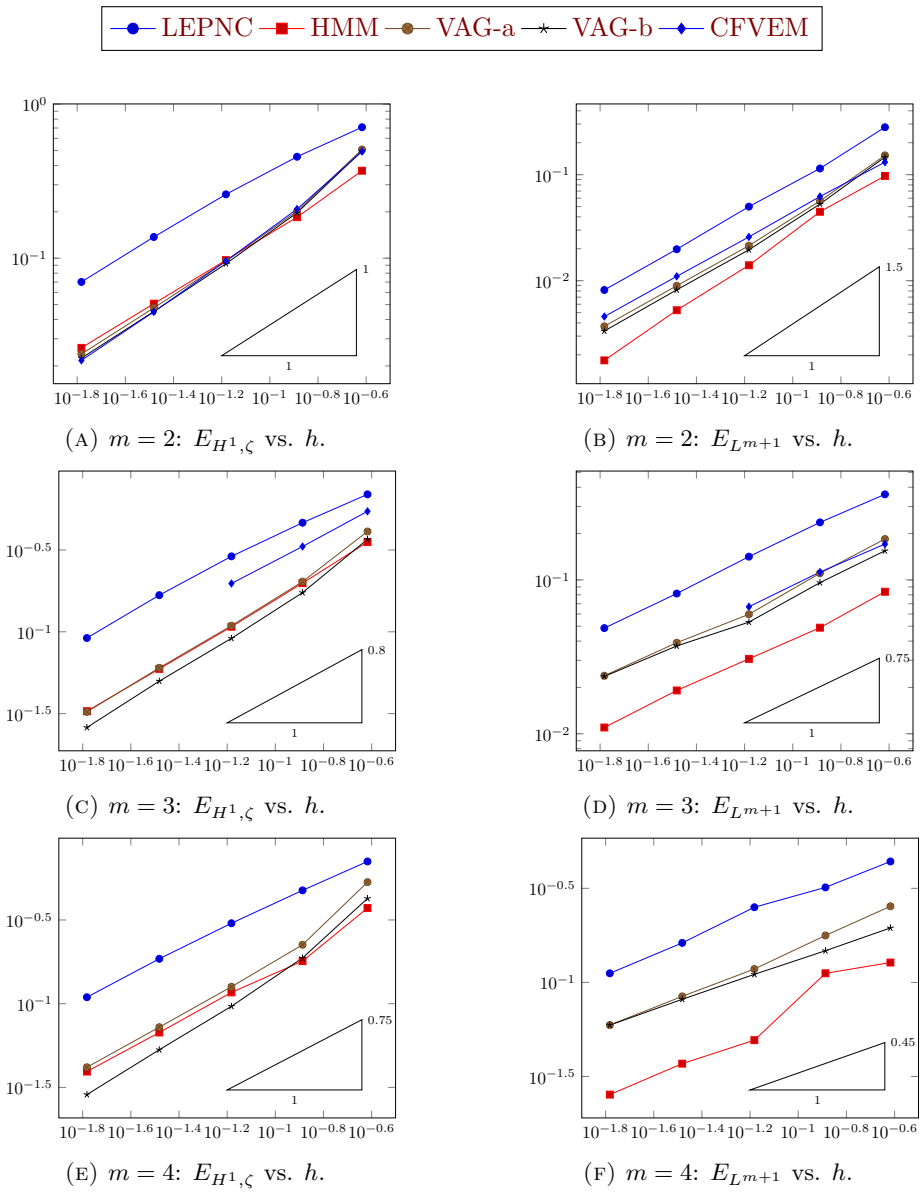
Comparing the various schemes, they all seem to adopt similar rates of convergence for $E_{H^1,\zeta}$ on hexagonal and locally refined Cartesian meshes; the differences mostly lie in the multiplicative constants, with the largest factor between these multiplicative constants of order 10. More variation in the rates is observed on these mesh families for $E_{L^{m+1}}$, which is probably due to the variation of m and reduced regularity of $\overline{u}$, as discussed above. The rates on triangular meshes seem to depend much more on the chosen scheme. Focusing on $E_{H^1,\zeta}$, which is a more stable measure, we see that MLP1 outperforms the other schemes, at least for $m = 2, 3$; of course, the drawback of MLP1 is that it can only be applied on triangular meshes. The other outlier is HDG, whose rates are much lower than the other schemes; the reason for that might be found in the total number of degrees of freedom, after static condensation, which is lower for HDG than some other schemes (see discussion in Section 4.1.2), and which therefore prevents this scheme from achieving optimal rates with respect to the mesh size.

It can also be noticed that some schemes produce a better $E_{H^1,\zeta}$ error than others, but that the "ranking" between the schemes can be reversed if we look at the error $E_{L^{m+1}}$.

4.1.2. *Algebraic complexity.* We now briefly discuss the performance of the schemes relative to their algebraic complexity, measured primarily here in terms of their number `NDOFs` of degrees of freedom. Focusing only on $m = 4$ (the most severe case), and hexagonal and triangular grids, we plot in Figure 5 the energy error $E_{H^1,\zeta}$ of each scheme versus its degrees of freedom.

A first remark is that this measure is slightly more favourable to HDG than in the previous section. Its rate remains a bit lower than the other schemes, at least for the considered meshes, but perhaps less so than when comparing the error versus the mesh size. We also notice that, on triangular meshes, the (mostly) vertex-centered methods VAG-a, VAG-b, and MLP1 outperform the other schemes, which is expected: on triangular meshes, vertex-based methods have much fewer degrees of freedom than edge-based methods such as LENCP or HMM. Curiously, using the same argument, we would expect CFVEM, which is also a vertex-based method, to behave better than it does; this could be explained by the different kind of mass-lumping applied for these two methods.

On hexagonal meshes, except for LEPNC, all schemes presented here have a comparable error vs. complexity. The advantage of vertex-based methods is less perceptible on such meshes than on



(A) $m = 2$: $E_{H^1,\zeta}$ vs. h.

(B) $m = 2$: $E_{L^{m+1}}$ vs. h.

(C) $m = 3$: $E_{H^1,\zeta}$ vs. h.

(D) $m = 3$: $E_{L^{m+1}}$ vs. h.

(E) $m = 4$: $E_{H^1,\zeta}$ vs. h.

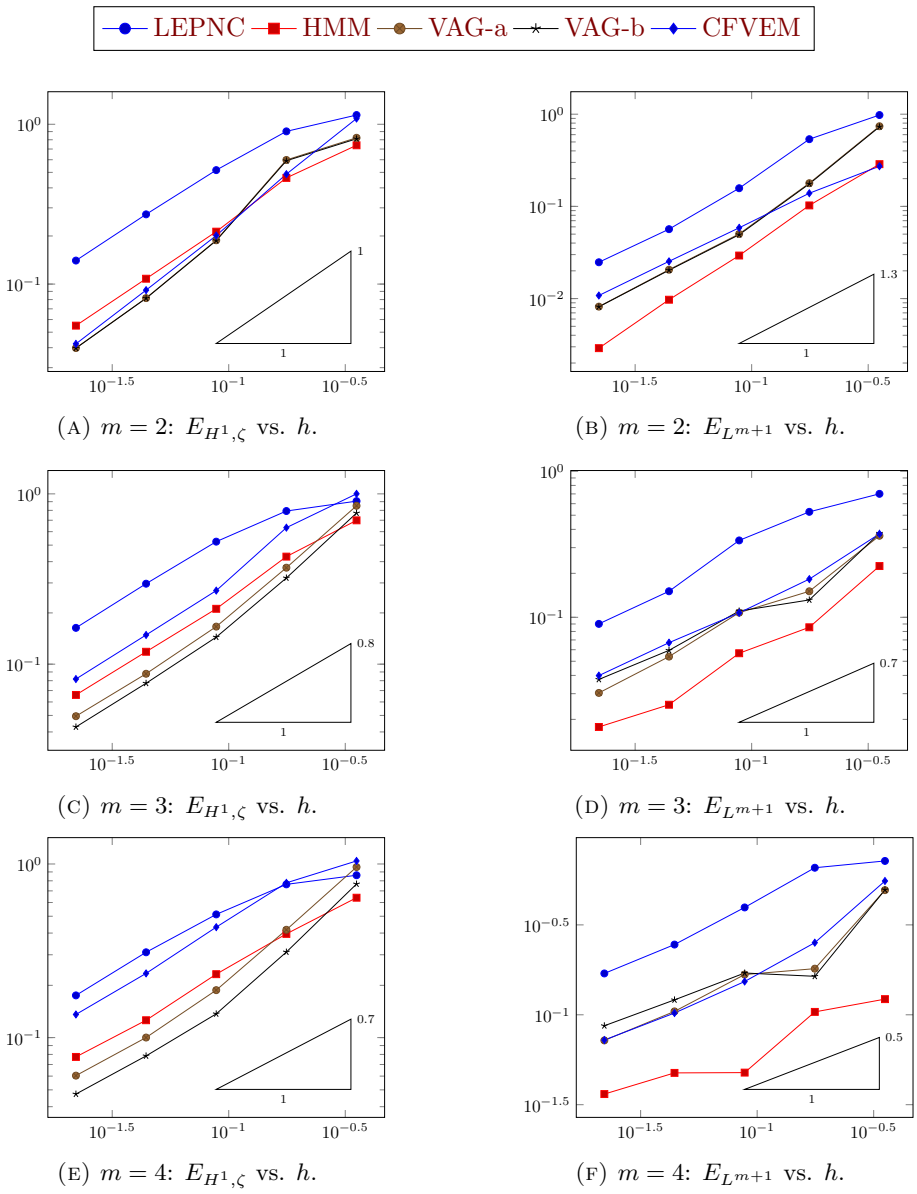
(F) $m = 4$: $E_{L^{m+1}}$ vs. h.

FIGURE 2. Hexagonal meshes: errors versus mesh size.

triangular meshes, which is not surprising: triangular meshes roughly have three times more edges than vertices, while hexagonal meshes have 1.5 times more edges than vertices on average.

4.1.3. *Positivity.* Finally, we look at the positivity properties of the schemes. As the standard linear heat equation, the porous medium equation satisfies a maximum principle: if the initial solution and the source terms are positive, then the solution remains positive for all times. Maintaining this property at the discrete level is particularly challenging, especially for schemes designed for generic polygonal/polyhedral meshes [20]. Except for MLP1 (which is restricted to triangular meshes), none of the schemes presented here satisfy this property in general.

In Figures 6, 7 and 8, we present the log-log graphs of the relative negative masses `NMass` versus the mesh sizes. In most situations, the schemes produce some negative mass, but it decays as the mesh is refined and is rather small relative to the total mass of the solution at the final time. On hexagonal and locally refined Cartesian meshes, the VAG and CFVEM schemes –which are (mostly) vertex-centered– have better positivity properties than the edge-centered schemes LEPNC and HMM, with VAG-b outperforming all the other schemes. Surprisingly, perhaps, VAG-a and VAG-b do not even produce negative values at the final time on locally refined meshes (and are thus absent from Figure 7). On triangular meshes, however, VAG-a produces more negative mass



(A) $m = 2$: $E_{H^1,\zeta}$ vs. h.

(B) $m = 2$: $E_{L^{m+1}}$ vs. h.

(C) $m = 3$: $E_{H^1,\zeta}$ vs. h.

(D) $m = 3$: $E_{L^{m+1}}$ vs. h.

(E) $m = 4$: $E_{H^1,\zeta}$ vs. h.

(F) $m = 4$: $E_{L^{m+1}}$ vs. h.

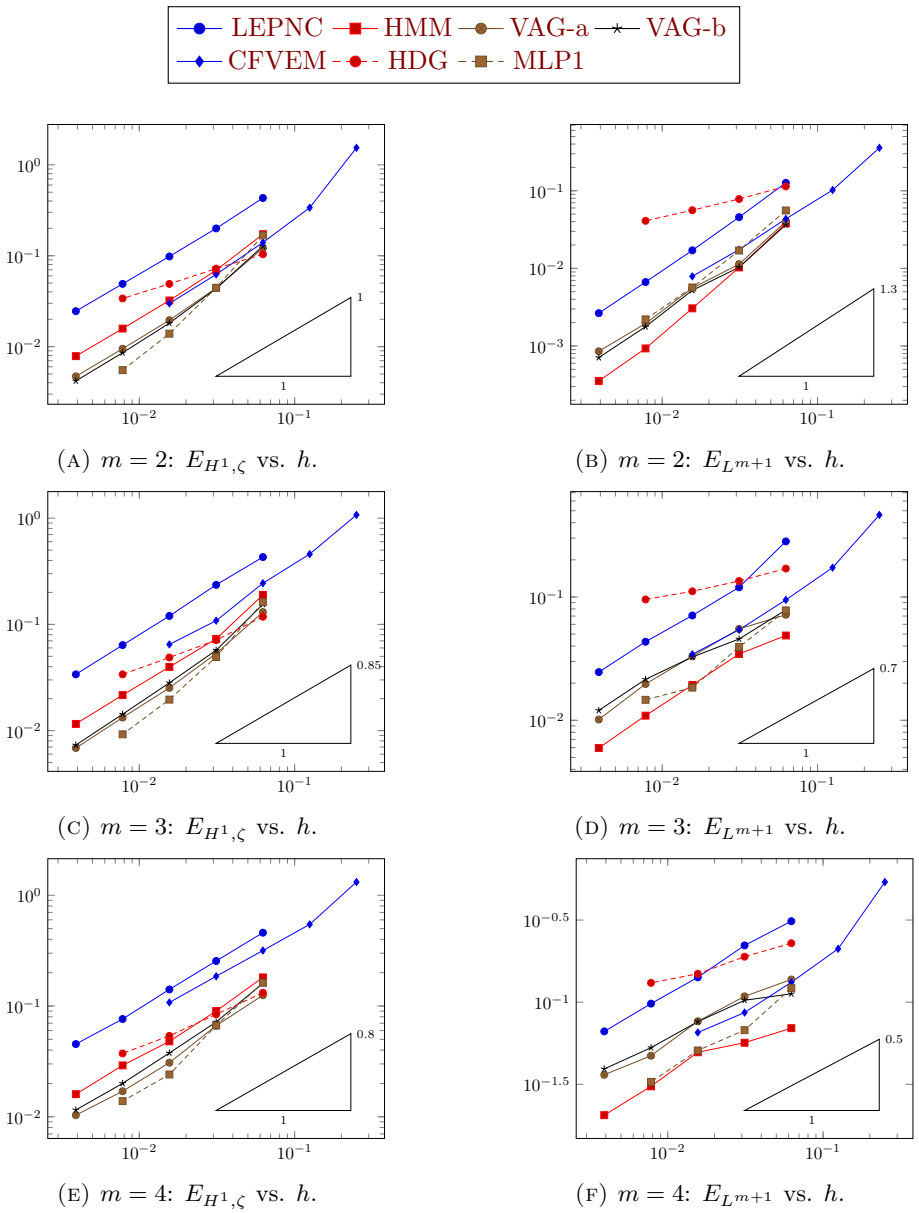
FIGURE 3. Locally refined Cartesian meshes: errors versus mesh size.

than the other schemes (and an amount comparable to HDG), with HMM and VAG-b being much closer to each other (with relative performance depending on m), and LEPNC in between these clusters. MLP1 is known to preserve positive values and is therefore not represented in Figure 8. CFVEM was also found to preserve positive values on these triangular meshes.

These results demonstrate a strong interaction between scheme design and mesh geometries when it comes to preserving the maximum principle of the continuous model.

4.2. **A word on non-linear iterative schemes.** The time discrete problems (7) or the fully discrete counterparts (30) are nonlinear. To determine an approximation of the solution one needs to employ an iterative method. The common choice is the Newton method (see, e.g., [8]), which converges quadratically. However, this convergence is guaranteed only if the initial guess is close enough to the solution. A choice at hand being the solution computed at the previous time step, this means that the convergence is guaranteed if the differences between the solutions at two successive times are small enough. This induces restrictions on the time step.

For (1), the Newton method can fail to converge due to the singularities of ζ, and in particular, for the fast diffusion case. To overcome this, one can regularise ζ to avoid degeneracy, but even in this case, the convergence is only guaranteed under severe restrictions on the time step. To



(A) $m = 2$: $E_{H^1,\zeta}$ vs. h. (B) $m = 2$: $E_{L^{m+1}}$ vs. h.

(C) $m = 3$: $E_{H^1,\zeta}$ vs. h. (D) $m = 3$: $E_{L^{m+1}}$ vs. h.

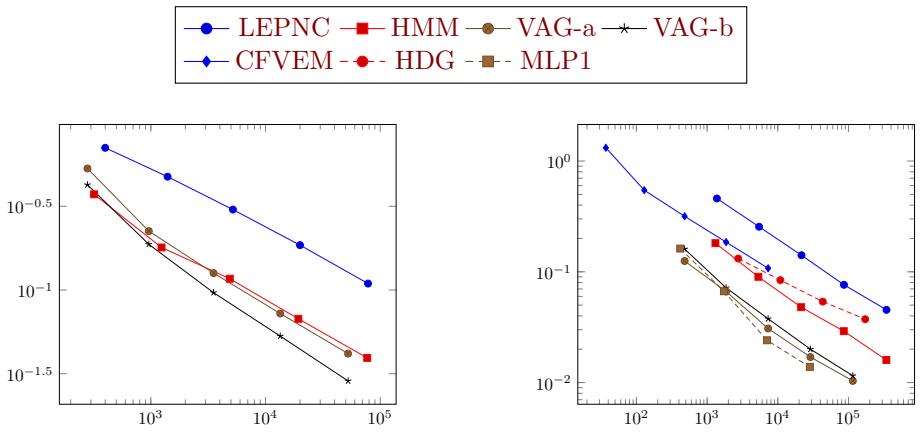
(E) $m = 4$: $E_{H^1,\zeta}$ vs. h. (F) $m = 4$: $E_{L^{m+1}}$ vs. h.

FIGURE 4. Triangular meshes: errors versus mesh size.

address these shortcomings, alternative iterative schemes have been designed. We mention here the relaxation scheme in [33], which shows to be more stable w.r.t. the choice of the initial condition, and the modified Picard scheme in [16], which is a simplified version of the Newton method. Both schemes are converging linearly. For these, as for the Newton scheme, the convergence is guaranteed rigorously under severe restrictions for the time step, as proved in [47].

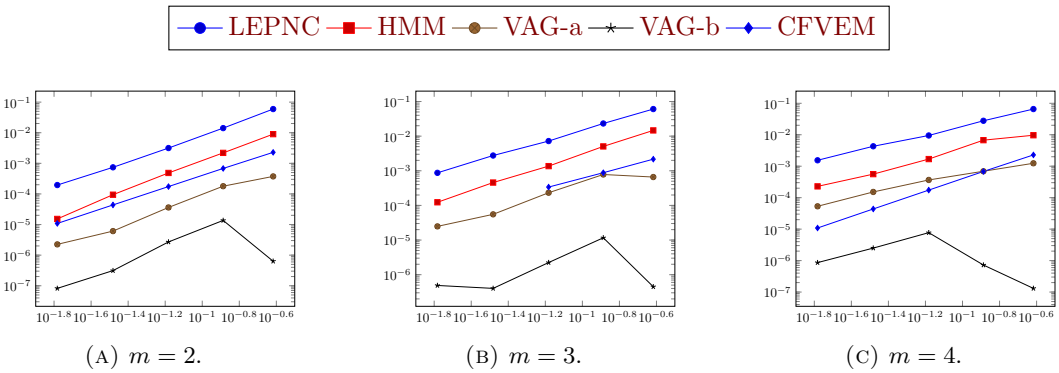
A fixed point (contraction) scheme exploiting the monotonicity of ζ has been proposed in [43] for the fast diffusion case and extended to more general situations in [46]. Though linear, the convergence is guaranteed under mild restrictions on the time step, regardless of the initial guess, and for any spatial discretisation. Moreover, as shown in [37], this scheme can be used to obtain a good initial guess for the Newton scheme, which leads to a stable and fast convergent iterative method. We also mention the scheme in [40], where the fixed point approach is combined with the Picard or Newton method by adding a is stabilisation term. This leads to a scheme with the stability of the fixed point scheme and converging like the Picard scheme.

Finally, we refer to [9], where both $\overline{u}$ and $\zeta(\overline{u})$ are expressed in terms of a different unknown, based on a properly chosen parametrisation. This allows reformulating the Newton method in such



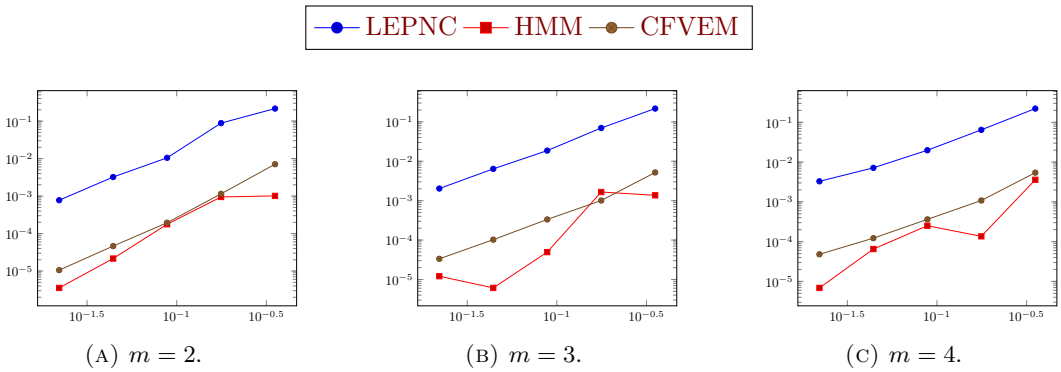
(A) Hexagonal meshes. (B) Triangular meshes.

FIGURE 5. Energy error $E_{H^1,\zeta}$ versus number of degrees of freedom `NDOFs`, $m = 4$.



(A) $m = 2$. (B) $m = 3$. (C) $m = 4$.

FIGURE 6. Hexagonal meshes: fraction of negative mass `NMass` versus mesh size.

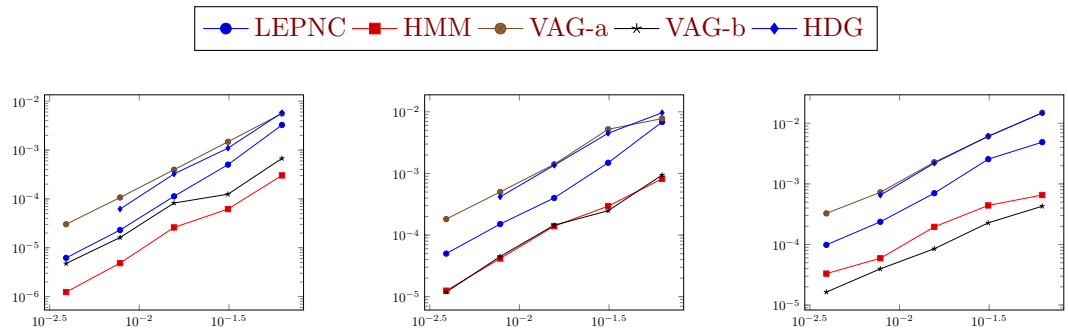


(A) $m = 2$. (B) $m = 3$. (C) $m = 4$.

FIGURE 7. Locally refined Cartesian meshes: fraction of negative mass `NMass` versus mesh size.

a way that the quadratic convergence is unaffected, but the stability is significantly improved, so that larger time steps are allowed.

Here we discuss a simple iterative scheme that is inspired by the fixed point approach in [37, 43]. We restrict to the case where ζ is Lipschitz continuous and let L_ζ denote the Lipschitz constant, but the idea can be extended to more general situations as well, e.g., by applying the idea in [9]. For the ease of presentation, we present the scheme in the time-discrete case, the fully discrete one being analogous. With $n \in \{0, \dots, N-1\}$ fixed we start by observing that (7) can be rewritten as



(A) $m = 2$. (B) $m = 3$. (C) $m = 4$.

FIGURE 8. Triangular meshes: fraction of negative mass `NMass` versus mesh size.

the system

$$\begin{aligned}(\overline{u}^{(n+1)},\varphi)+\delta t^{(n+\frac12)}(\nabla\overline{w}^{(n+1)},\nabla\varphi)&=(\overline{u}^{(n)},\varphi)+\delta t^{(n+\frac12)}(f^{(n+1)},\varphi),\\(\overline{w}^{(n+1)},\psi)&=(\zeta(\overline{u}^{(n+1)}),\psi),\end{aligned}\tag{61}$$

for all $\varphi\in H^1_0(\Omega)$ and $\psi\in L^2(\Omega)$. For a given $L\geq\frac{L_\zeta}{2}$, the iterative scheme consists in finding the pairs $(\overline{u}^i,\overline{w}^i)\in H^1_0(\Omega)\times L^2(\Omega)$ $(i\in\mathbb{N}^*)$ solving the linear systems

$$\begin{aligned}(\overline{u}^i,\varphi)+\delta t^{(n+\frac12)}(\nabla\overline{w}^i,\nabla\varphi)&=(\overline{u}^{(n)},\varphi)+\delta t^{(n+\frac12)}(f^{(n+1)},\varphi),\\(\overline{w}^i,\psi)&=L(\overline{u}^i-\overline{u}^{i-1},\psi)+(\zeta(\overline{u}^{i-1}),\psi),\end{aligned}\tag{62}$$

for all $\varphi\in H^1_0(\Omega)$ and $\psi\in L^2(\Omega)$. A natural choice for the initial guess is $\overline{u}^0=\overline{u}^{(n)}$ (the solution at the previous time step), but, as will be seen below, the convergence is guaranteed for any starting point. To prove this, we define the iteration errors

$$e^i_u=\overline{u}^{(n+1)}-\overline{u}^i,\qquad\text{and}\qquad e^i_w=\zeta(\overline{u}^{(n+1)})-\overline{w}^i.\tag{63}$$

From (61) and (62), the errors satisfy

$$\begin{aligned}(e^i_u,\varphi)+\delta t^{(n+\frac12)}(\nabla e^i_w,\nabla\varphi)&=0,\\(e^i_w,\psi)&=L(e^i_u-e^{i-1}_u,\psi)+(\zeta(\overline{u}^{(n+1)})-\zeta(\overline{u}^{i-1}),\psi),\end{aligned}\tag{64}$$

for all $\varphi\in H^1_0(\Omega)$ and $\psi\in L^2(\Omega)$. With this, the convergence result is

Lemma 4.2. *The iterative scheme in (62) is convergent regardless of the initial guess. More precisely, one has $\overline{w}^i\to\zeta(\overline{u}^{(n+1)})$ in $H^1(\Omega)$ and $\overline{u}^i\to\overline{u}^{(n+1)}$ in $L^2(\Omega)$ as $i\to\infty$.*

Proof. Taking $\varphi=e^i_w$ and $\psi=e^i_u$ into (64) and subtracting the resulting gives

$$L\|e^i_u\|^2+\delta t^{(n+\frac12)}\|\nabla e^i_w\|^2=(Le^{i-1}_u-(\zeta(\overline{u}^{(n+1)})-\zeta(\overline{u}^{i-1})),e^i_u).$$

Since ζ is Lipschitz, by the choice of L one has $|Le^{i-1}_u-(\zeta(\overline{u}^{(n+1)})-\zeta(\overline{u}^{i-1}))|\leq L|e^{i-1}_u|$. This, together with the Cauchy–Schwarz inequality leads to

$$L\|e^i_u\|^2+\delta t^{(n+\frac12)}\|\nabla e^i_w\|^2\leq L\|e^{i-1}_u\|\|e^i_u\|.$$

Applying now (12) and multiplying the resulting by 2 yields

$$L\|e^i_u\|^2+2\delta t^{(n+\frac12)}\|\nabla e^i_w\|^2\leq L\|e^{i-1}_u\|^2.\tag{65}$$

Adding (65) for $i=1,\dots,k$ (k being arbitrary) leads to

$$L\|e^k_u\|^2+2\delta t^{(n+\frac12)}\sum_{i=1}^k\|\nabla e^i_w\|^2\leq L\|e^0_u\|^2.\tag{66}$$

This shows that the second term above is a convergent series, implying the first convergence result. Using this convergence in the first equation of (64) completes the proof. □

Acknowledgments

CC acknowledges support from the Labex CEMPI (ANR-11-LABX-0007-01). JD and GM were partially supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP170100605); GM was also partially supported by the ERC Project CHANGE, which has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 694515). MBO and ISP are supported by the the Research Foundation-Flanders (FWO), Belgium through the Odysseus programme project G0G1316N.

References

[1] HArDCore2D – Hybrid Arbitrary Degree::Core 2D. https://github.com/jdroniou/HArDCore2D-release, Version 2.0.2.
[2] H. W. Alt and S. Luckhaus. Quasilinear elliptic-parabolic differential equations. *Math. Z.*, 183:311–341, 1983.
[3] B. Andreianov, C. Cancès, and A. Moussa. A nonlinear time compactness result and applications to discretization of degenerate parabolic-elliptic PDEs. *J. Funct. Anal.*, 273(12):3633–3670, 2017.
[4] O. Angelini, K. Brenner, and D. Hilhorst. A finite volume method on general meshes for a degenerate parabolic convection-reaction-diffusion equation. *Numer. Math.*, 123(2):219–257, 2013.
[5] T. Arbogast, M. F. Wheeler, and N.-Y. Zhang. A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. *SIAM J. Numer. Anal.*, 33(4):1669–1687, 1996.
[6] J. W. Barrett and P. Knabner. Finite element approximation of the transport of reactive solutes in porous media. II. Error estimates for equilibrium adsorption processes. *SIAM J. Numer. Anal.*, 34(2):455–479, 1997.
[7] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo. Basic principles of virtual element methods. *Math. Models Methods Appl. Sci. (M3AS)*, 199(23):199–214, 2013.
[8] L. Bergamaschi and M. Putti. Mixed finite elements and Newton-type linearizations for the solution of Richards' equation. *Internat. J. Numer. Methods Engrg.*, 45(8):1025–1046, 1999.
[9] K. Brenner and C. Cancès. Improving Newton's method performance by parametrization: the case of the Richards equation. *SIAM J. Numer. Anal.*, 55(4):1760–1785, 2017.
[10] C. Cances. Energy stable numerical methods for porous media flow type problems. *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles*, 73, 2018.
[11] C. Cancès and T. Gallouët. On the time continuity of entropy solutions. *J. Evol. Equ.*, 11(1):43–55, 2011.
[12] C. Cancès and C. Guichard. Convergence of a nonlinear entropy diminishing Control Volume Finite Element scheme for solving anisotropic degenerate parabolic equations. *Math. Comp.*, 85(298):549–580, 2016.
[13] C. Cancès and C. Guichard. Numerical analysis of a robust free energy diminishing finite volume scheme for parabolic equations with gradient structure. *Found. Comput. Math.*, 17(6):1525–1584, 2017.
[14] C. Cancès, I. S. Pop, and M. Vohralík. An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow. *Math. Comp.*, 83(285):153–188, 2014.
[15] J. Carrillo. Entropy solutions for nonlinear degenerate problems. *Arch. Ration. Mech. Anal.*, 147(4):269–361, 1999.
[16] M. Celia, E. Bouloutas, and R. Zarba. A General Mass-Conservative Numerical-Soution for the Unsaturated Flow Equation. *Water Resour. Res.*, 26(7):1483–1496, 1990.
[17] B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47(2):1319–1365, 2009.
[18] D. A. Di Pietro and J. Droniou. A third Strang lemma and an Aubin-Nitsche trick for schemes in fully discrete formulation. *Calcolo*, 55(3):Art. 40, 39p, 2018.
[19] D. A. Di Pietro and J. Droniou. *The Hybrid High-Order Method for Polytopal Meshes: Design, Analysis, and Applications*, volume 19 of *Modeling, Simulation and Applications*. Springer International Publishing, 2020.
[20] J. Droniou. Finite volume schemes for diffusion equations: introduction to and review of modern methods. *Math. Models Methods Appl. Sci. (M3AS)*, 24(8):1575–1619, 2014. Special issue on Recent Techniques for PDE Discretizations on Polyhedral Meshes.
[21] J. Droniou and R. Eymard. Uniform-in-time convergence of numerical methods for non-linear degenerate parabolic equations. *Numer. Math.*, 132(4):721–766, 2016.
[22] J. Droniou and R. Eymard. High-order mass-lumped schemes for nonlinear degenerate elliptic equations. *SIAM J. Numer. Anal.*, 58(1):153–188, 2020.
[23] J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. *The gradient discretisation method*, volume 82 of *Mathematics & Applications*. Springer, 2018.
[24] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. Non-conforming finite elements on polytopal meshes.
[25] J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. Gradient schemes: a generic framework for the discretisation of linear, nonlinear and nonlocal elliptic and parabolic equations. *Math. Models Methods Appl. Sci. (M3AS)*, 23(13):2395–2432, 2013.
[26] Y. Epshteyn and B. Rivière. Analysis of *hp* discontinuous Galerkin methods for incompressible two-phase flow. *J. Comput. Appl. Math.*, 225(2):487–509, 2009.
[27] A. Ern and I. Mozolevski. Discontinuous Galerkin method for two-component liquid-gas porous media flows. *Comput. Geosci.*, 16(3):677–690, 2012.
[28] R. Eymard, P. Féron, T. Gallouët, C. Guichard, and R. Herbin. Gradient schemes for the Stefan problem. *Int. J. Finite Vol.*, 13:1–37, 2013.

[29] R. Eymard, T. Gallouët, D. Hilhorst, and Y. Naït Slimane. Finite volumes and nonlinear diffusion equations. *RAIRO Modél. Math. Anal. Numér.*, 32(6):747–761, 1998.
[30] R. Eymard, R. Herbin, and A. Michel. Mathematical study of a petroleum-engineering scheme. *M2AN Math. Model. Numer. Anal.*, 37(6):937–972, 2003.
[31] R. Eymard, D. Hilhorst, and M. Vohralík. A combined finite volume–nonconforming/mixed-hybrid finite element scheme for degenerate parabolic problems. *Numer. Math.*, 105(1):73–131, 2006.
[32] J. G. Heywood and R. Rannacher. Finite-element approximation of the nonstationary Navier-Stokes problem. IV. Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2):353–384, 1990.
[33] W. Jäger and J. Kačur. Solution of doubly nonlinear and degenerate parabolic problems by relaxation schemes. *RAIRO Modél. Math. Anal. Numér.*, 29(5):605–627, 1995.
[34] R. A. Klausen, F. A. Radu, and G. T. Eigestad. Convergence of MPFA on triangulations and for Richards' equation. *Internat. J. Numer. Methods Fluids*, 58(12):1327–1351, 2008.
[35] O. A. Ladyženskaja, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and quasilinear equations of parabolic type*. Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1967.
[36] N. Liao. A unified approach to the Holder regularity of solutions to degenerate and singular parabolic equations. *J. Differential Equations*, 268(10):5704–5750, 2020.
[37] F. List and F. A. Radu. A study on iterative methods for solving Richards' equation. *Comput. Geosci.*, 20(2):341–353, 2016.
[38] E. Magenes, R. H. Nochetto, and C. Verdi. Energy error estimates for a linear scheme to approximate nonlinear parabolic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 21(4):655–678, 1987.
[39] A. M. Meirmanov. *The Stefan problem*, volume 3 of *de Gruyter Expositions in Mathematics*. Walter de Gruyter & Co., Berlin, 1992. Translated from the Russian by Marek Niezgódka and Anna Crowley, With an appendix by the author and I. G. Götz.
[40] K. Mitra and I. S. Pop. A modified L-scheme to solve nonlinear diffusion problems. *Comput. Math. Appl.*, 77(6):1722–1738, 2019.
[41] R. H. Nochetto and C. Verdi. Approximation of degenerate parabolic problems using numerical integration. *SIAM J. Numer. Anal.*, 25(4):784–814, 1988.
[42] F. Otto. L^1-contraction and uniqueness for quasilinear elliptic-parabolic equations. *J. Differ. Equations*, 131:20–38, 1996.
[43] I. S. Pop, F. Radu, and P. Knabner. Mixed finite elements for the Richards' equation: linearization procedure. *J. Comput. Appl. Math.*, 168(1-2):365–373, 2004.
[44] I. S. Pop and B. Schweizer. Regularization schemes for degenerate Richards equations and outflow conditions. *Math. Models Methods Appl. Sci.*, 21(8):1685–1712, 2011.
[45] I. S. Pop, M. Sepúlveda, F. A. Radu, and O. P. Vera Villagrán. Error estimates for the finite volume discretization for the porous medium equation. *J. Comput. Appl. Math.*, 234(7):2135–2142, 2010.
[46] F. A. Radu, K. Kumar, J. M. Nordbotten, and I. S. Pop. A robust, mass conservative scheme for two-phase flow in porous media including Hölder continuous nonlinearities. *IMA J. Numer. Anal.*, 38(2):884–920, 2018.
[47] F. A. Radu, I. S. Pop, and P. Knabner. Newton-type methods for the mixed finite element discretization of some degenerate parabolic equations. In *Numerical mathematics and advanced applications*, pages 1192–1200. Springer, Berlin, 2006.
[48] F. A. Radu, I. S. Pop, and P. Knabner. Error estimates for a mixed finite element discretization of some degenerate parabolic equations. *Numer. Math.*, 109(2):285–311, 2008.
[49] J. L. Vázquez. *Smoothing and decay estimates for nonlinear diffusion equations*, volume 33 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2006. Equations of porous medium type.
[50] J. L. Vázquez. *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford, 2007. Mathematical theory.
[51] M. Vohralík and M. F. Wheeler. A posteriori error estimates, stopping criteria, and adaptivity for two-phase flows. *Comput. Geosci.*, 17(5):789–812, 2013.
[52] C. S. Woodward and C. N. Dawson. Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media. *SIAM J. Numer. Anal.*, 37(3):701–724, 2000.
[53] I. Yotov. A mixed finite element discretization on non-matching multiblock grids for a degenerate parabolic equation arising in porous media flow. *East-West J. Numer. Math.*, 5(3):211–230, 1997.
[54] W. P. Ziemer. Interior and boundary continuity of weak solutions of degenerate parabolic equations. *Trans. Amer. Math. Soc.*, 271:733–748, 1982.

Clément Cancès (clement.cances@inria.fr): Inria, Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille.

Jérôme Droniou (jerome.droniou@monash.edu): School of Mathematics, Monash University, Melbourne, Australia.

Cindy Guichard (cindy.guichard@sorbonne-universite.fr): Sorbonne Université, CNRS, Université de Paris, INRIA ANGE, Laboratoire Jacques-Louis Lions, F-75005 Paris, France.

Gianmarco Manzini (marco.manzini@imati.cnr.it): Istituto di Matematica Applicata e Tecnologie Informatiche - CNR, via Ferrata 1, Pavia, Italy.

Manuela Bastidas Olivares (manuela.bastidas@uhasselt.be): Hasselt University, Campus Diepenbeek, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium.

Iuliu Sorin Pop (sorin.pop@uhasselt.be): Hasselt University, Campus Diepenbeek, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium.